

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 19:57:20 ; Search time 2487.88 seconds
(without alignments)
7673.783 Million cell updates/sec

Title: US-09-581-500B-12
Perfect score: 656
Sequence: 1 gccacacacaaatgaat.....ccacgcgcagaccgcgga 656

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pla.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vit.*
- 38: em.sv.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	652.6	99.5	656	6	AX021006	AX021006 Sequence
c 2	647.8	98.8	191395	9	AC009802	AC009802 Homo sapi
c 3	646.2	98.5	198291	2	AC067875	AC067875 Homo sapi
c 4	646.2	98.5	201734	2	AC090224	AC090224 Homo sapi
c 5	643	98.0	191793	2	AP001897	AP001897 Homo sapi
c 6	441.6	67.3	13202	6	AX346386	AX346386 Sequence
c 7	376	57.3	186351	2	AP001503	AP001503 Homo sapi
c 8	367.8	56.1	13202	6	AX346387	AX346387 Sequence
c 9	73.4	11.2	1425	6	AR005195	AR005195 Sequence
c 10	73.4	11.2	1425	6	IR1228	IR1228 Sequence 1
c 11	73.4	11.2	1425	6	IR2208	IR2208 Sequence 1
c 12	73	11.1	1325	9	HUMCAP2A	I40377 Homo sapien
c 13	72	11.0	7218	6	I66494	I66494 Sequence 14
c 14	59.6	9.1	113684	2	AC103021	AC103021 Rattus no
c 15	57.4	8.8	65711	2	AC117583	AC117583 Mus muscu
c 16	56.8	8.7	750	11	PM12B11B	AL684309 Penicilli
c 17	55.4	8.4	130051	2	AC127840	AC127840 Rattus no
c 18	55.2	8.4	161310	2	AC120603	AC120603 Rattus no
c 19	54.8	8.4	156782	2	AC114101	AC114101 Rattus no
c 20	54	8.2	168542	2	AC058792	AC058792 Homo sapi
c 21	54	8.2	189218	9	AL355297	AL355297 Human DNA
c 22	53.8	8.2	83862	2	AC120671	AC120671 Rattus no
c 23	53.6	8.2	6170	9	HSCGSE7	AL672255 Mus muscu
c 24	53.6	8.2	169232	2	AC129646	AC129646 Rattus no
c 25	53.6	8.2	245249	2	AC103134	AC103134 Homo sapi
c 26	53.4	8.1	103874	2	AC106598	AC106598 Rattus no
c 27	53.4	8.1	194260	2	AP005508	AP005508 Oryza sat
c 28	53	8.1	7450	9	AB035726	AB035726 Homo sapi
c 29	53	8.1	145489	2	AL672255	AL672255 Mus muscu
c 30	53	8.1	220201	10	AL669898	AL669898 Mouse DNA
c 31	52.8	8.0	137721	2	AC113928	AC113928 Rattus no
c 32	52.6	8.0	119825	2	AC128152	AC128152 Rattus no
c 33	52.6	8.0	228584	2	AL669838	AL669838 Mus muscu
c 34	52.4	8.0	150653	2	AC109865	AC109865 Rattus no
c 35	52	7.9	922	9	HS034984	U92984 Homo sapien
c 36	52	7.9	1319	9	BC034528	BC034528 Homo sapi
c 37	52	7.9	193446	2	AC094597	AC094597 Rattus no
c 38	51.8	7.9	229380	2	AC079636	AC079636 Mus muscu
c 39	51.6	7.9	125020	9	AF429315	AF429315 Homo sapi
c 40	51.6	7.9	139838	2	AC119485	AC119485 Rattus no
c 41	51.4	7.8	106039	2	AC105476	AC105476 Rattus no
c 42	51.4	7.8	150074	2	AC016007	AC016007 Homo sapi
c 43	51.4	7.8	151729	2	AC124866	AC124866 Rattus no
c 44	51	7.8	76118	2	AC094984	AC094984 Rattus no
c 45	50.6	7.7	259	9	AB018491	AB018491 Homo sapi

ALIGNMENTS

RESULT 1
AX021006
LOCUS AX021006 656 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 12 from Patent WO9932643.
ACCESSION AX021006
VERSION AX021006.1 GI:10044669
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 656)
AUTHORS Del-Favero,J., Raeymaekers,P. and Van Broeckhoven,C.
TITLE Mood disorder gene
JOURNAL Patent: WO 9932643-A 12 01-JUL-1999;

BROECKHOVEN CHRISTINE VAN (BE); DEL FAVERO JURGEN (BE); RAEYMARKERS
PETER (BE); VIAAMS INTERUNIV INST BIOTECH (BE)

FEATURES
source
Location/Qualifiers
1. .656
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT
ORIGIN
183 a 165 c 208 g 96 t 4 others

Query Match 99.5%; Score 652.6; DB 6; Length 656;
Best Local Similarity 100.0%; Pred. No. 8.1e-148; Indels 0; Gaps 0;
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCACAAACAAATGAATGAACCTGGGATGATTTTGGCCAGCAATAGAAAA 60
Db 1 GCACAAACAAATGAATGAACCTGGGATGATTTTGGCCAGCAATAGAAAA 60
QY 61 TGATTAGTATCCTTATCAGGAGCAATTCAGAGATGTTGGTGGACCTCAACTACAG 120
Db 61 TGATTAGTATCCTTATCAGGAGCAATTCAGAGATGTTGGTGGACCTCAACTACAG 120
QY 121 TGGAGTCAACGCTGAATCAACGCTGAAAGAGCAATAGCAATGTGTACACTTTTAT 180
Db 121 TGGAGTCAACGCTGAATCAACGCTGAAAGAGCAATAGCAATGTGTACACTTTTAT 180
QY 181 AAAAACCACCTCCAGGACCCAGGACTGGCCCTCTCTCCGGTGCCACAGACATCCACA 240
Db 181 AAAAACCACCTCCAGGACCCAGGACTGGCCCTCTCTCCGGTGCCACAGACATCCACA 240
QY 241 CAGGCCCCAAATCAAGGATTCACAGCCAGACCAATCAAGCGTTCTGTGATCATCTG 300
Db 241 CAGGCCCCAAATCAAGGATTCACAGCCAGACCAATCAAGCGTTCTGTGATCATCTG 300
QY 301 CCGAGAGCTTGCCTCAATCAAGCGGACGTAAGCAATCTACAAAGGAGGATAGTCAA 360
Db 301 CCGAGAGCTTGCCTCAATCAAGCGGACGTAAGCAATCTACAAAGGAGGATAGTCAA 360
QY 361 AGCAGCAGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCT 420
Db 361 AGCAGCAGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCT 420
QY 421 CAGAGTACCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCT 480
Db 421 CAGAGTACCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCT 480
QY 481 CTGGAGTGGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540
Db 481 CTGGAGTGGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
QY 541 GCGGGCAGAACTANGAATGACSCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
Db 541 GCGGGCAGAACTANGAATGACSCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
QY 601 GTGGGGTCCCTGGCAGGCGGGGTTCCCAAGCCCGGAGGAGGAGGAGGAGGAGGAG 656
Db 601 GTGGGGTCCCTGGCAGGCGGGGTTCCCAAGCCCGGAGGAGGAGGAGGAGGAGGAG 656

RESULT 2
AC009802/c
LOCUS AC009802 191395 bp DNA linear PRI 17-FEB-2000
DEFINITION Homo sapiens chromosome 18, clone RP11-793J2, complete sequence.
ACCESSION AC009802
VERSION AC009802.13 GI:6705901
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 191395)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-793J2
JOURNAL Unpublished

REFERENCE
AUTHORS

2 (bases 1 to 191395)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funks,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagou,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tessaye,S., Torruella-Miller,I., Vassiliev,H., Vo.A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

TITLE
JOURNAL

Submitted (01-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 191395)

REFERENCE
AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepl,X., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Dominc,M., Doyle,M.,
Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagou,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,S., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tessaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triggilio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

TITLE
JOURNAL

Direct Submission
Submitted (17-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 16, 2000 this sequence version replaced gi:6514043.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Center project name: L1250
Center clone name: 793_J_2

FEATURES
source

Location/Qualifiers
1. .191395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-793J2"
/clone_lib="RPCI-11 Human Male BAC"
2. .542
/rpt_family="L1M2"
repeat_region 474..708
repeat_region /rpt_family="L1"
repeat_region 708..1353
repeat_region /rpt_family="L1"
repeat_region 1351..1623
repeat_region /rpt_family="L1MA4A"

repeat_region	1632..1687
unsure	/rpt_family="TATATG)n" complement(1673..1726)
repeat_region	/note="Single-stranded coverage." 1688..2207
unsure	/rpt_family="L1MA4A" complement(1724..1770)
repeat_region	/note="Single-stranded terminator coverage." 2208..2215
repeat_region	/rpt_family="TATATG)n" 2594..2648
repeat_region	/rpt_family="L1ME" complement(2649..2991)
repeat_region	/rpt_family="MER2" 2992..3475
repeat_region	/rpt_family="L1MB" complement(4330..4437)
repeat_region	/rpt_family="FRAM" 4438..4567
repeat_region	/rpt_family="(TA)n" 4599..4959
repeat_region	/rpt_family="AT_rich" 4982..5151
repeat_region	/rpt_family="(TATG)n" 5144..5311
repeat_region	/rpt_family="(TATAA)n" 5357..5441
repeat_region	/rpt_family="(TA)n" 5444..5525
repeat_region	/rpt_family="AT_rich" 5533..5708
repeat_region	/rpt_family="(TTATA)n" 5742..5907
repeat_region	/rpt_family="(TTATA)n" complement(5966..6047)
repeat_region	/rpt_family="FRAM/FAM" 6682..6703
repeat_region	/rpt_family="AT_rich" 7034..7933
repeat_region	/rpt_family="L1PA16" 7935..8006
repeat_region	/rpt_family="AluY" 8009..9111
repeat_region	/rpt_family="L1PA16" 9112..9165
repeat_region	/rpt_family="AT_rich" 9216..9496
repeat_region	/rpt_family="AluSx" 9497..9520
repeat_region	/rpt_family="(CAA)n" 9961..10264
unsure	/rpt_family="AluY" complement(10837..10840)
repeat_region	/note="Single-stranded terminator coverage." complement(11156..11515)
repeat_region	/rpt_family="HEIC" 11836..11987
repeat_region	/rpt_family="MTR" 12592..12643
repeat_region	/rpt_family="I2" 13683..13994
repeat_region	/rpt_family="AluSx" complement(13996..14160)
repeat_region	/rpt_family="LINEc" 14788..15255
repeat_region	/rpt_family="MER67D" complement(15875..15996)
repeat_region	/rpt_family="MERSB" 16101..16135
repeat_region	/rpt_family="AT_rich" 16136..16304
repeat_region	/rpt_family="L1M4" complement(16572..16647)

repeat_region	/rpt_family="MADEI" complement(16716, .17011)
repeat_region	/rpt_family="AluY" 17045, .17780
repeat_region	/rpt_family="LIME3" 17827, .17938
repeat_region	/rpt_family="LIME3" complement(17945, .18133)
repeat_region	/rpt_family="MER58A" 18222, .18755
repeat_region	/rpt_family="LIME3" 18843, .19019
repeat_region	/rpt_family="LIME3" 19573, .19616
repeat_region	/rpt_family="(CACG)n" 20506, .20544
repeat_region	/rpt_family="AT_rich" 21894, .22184
repeat_region	/rpt_family="AluSq" complement(22787, .22998)
repeat_region	/rpt_family="LFR40a" complement(23074, .23133)
repeat_region	/rpt_family="LFR40b" 23184, .23540
repeat_region	/rpt_family="L2" complement(23675, .25225)
repeat_region	/rpt_family="ptr5" 25329, .25582
repeat_region	/rpt_family="THE1B" 25879, .26031
repeat_region	/rpt_family="MLT1a1" 26032, .26321
repeat_region	/rpt_family="AluSx" 26322, .26531
repeat_region	/rpt_family="MLT1a1" 26823, .26736
repeat_region	/rpt_family="L2" 26943, .27439
repeat_region	/rpt_family="L2" 27314, .27985
repeat_region	/rpt_family="MIR" 27314, .27985

Query Match	98.8%;	Score 647.8;	DB 9;	Length 191395;
Best Local Similarity	98.9%;	Pred. No. 1.6e-146;		
Matches 649;	Conservative	1; Mismatches 6;	Indels 0;	Gaps
QY	1	GCACAACAACAAATGAATAGACACTGGGATGTATTTTGGCCAGGCAATTAGAATA	60	
Db	98274	GCACACAACAATGNATATAGACTGGGATGTATTTTGGCCAGGCAATTAGAATA	98215	
QY	61	TGATTAGTATCCTTATCAGAGAGCAATTTTCAGAAAGCTTTGGTGGAGCTTAACATACAG	120	
Db	98214	TGATTAGTATCCTTATCAGAGAGCAATTTTCAGAAATGTTTGGTGGAGCTTAACATACAG	98155	
QY	121	TGGAGTCAAAAGTCAATCAACAGGTGAAAAGAGGACAATGCCAAATGTGTACACTTTTAT	180	
Db	98154	TGGAGTCAAAAGTCAATCAACAGGTGAAAAGAGGACAATGCCAAATGTGTACACTTTTAT	98095	
QY	181	AAAAACCACCTCCACAGGACAGGCACATGGCCCTCTCTCGGTGCCACAGACATCCACA	240	
Db	98094	AAAAACCACCTCCACAGGACAGGCACATGGCCCTCTCTCGGTGCCACAGACATCCACA	98033	
QY	241	CAGGCCCAAGCAATTCAGGATTCACACAGCCAGAGCAATCGAACGGTTCAGTGATCTCG	300	
Db	98034	CAGGCCCAAGCAATTCAGGATTCACACAGCCAGAGCAATCGAACGGTTCAGTGATCTCG	97975	
QY	301	CGGAGGCGCTTGCCCTCAATCAAGCGGACGCTGAAGCATCTACAAGAGAGGAATAGTCAA	360	
Db	97974	CGGAGGCGCTTGCCCTCAATCAAGCGGACGCTGAAGCATCTACAAGAGAGGAATAGTCAA	97915	
QY	361	AGCAGACGGCGGCGGCGCGCGCGAGCAGCAGCAGCAGGAGGTGGGGCCCTCTG	420	
Db	97914	AGCAGACGGCGGCGGCGGCGCGCGAGCAGCAGCAGCAGGAGGTGGGGCCCTCTG	97855	


```

RESULT 4
AC090224/c
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-693A18 map 18, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
AC090224
VERSION
AC090224.4 GI:14190716
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 201734)
Biren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-693A18
Unpublished
2 (bases 1 to 201734)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camrata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,D., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., MacDonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Roqov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Triggilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 23, 2001 this sequence version replaced gi:13621272.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12265
Center clone name: 693_A_18
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 198542 bases at least Q40
Consensus quality: 200270 bases at least Q30
Consensus quality: 200829 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 201034; sum-of-contigs
Quality coverage: 11.3 in Q20 bases; agarose-fp
Quality coverage: 10.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```

```

* be preserved.
1 45183: contig of 45183 bp in length
* 45184 45283: gap of 100 bp
* 45284 46768: contig of 1485 bp in length
* 46769 46868: gap of 100 bp
* 46869 47961: contig of 1093 bp in length
* 47962 48061: gap of 100 bp
* 48062 49643: contig of 1582 bp in length
* 49644 49743: gap of 100 bp
* 49744 51393: contig of 1650 bp in length
* 51394 51493: gap of 100 bp
* 51494 74531: contig of 23038 bp in length
* 74532 74631: gap of 100 bp
* 74632 127710: contig of 53079 bp in length
* 127711 127810: gap of 100 bp
* 127811 201734: contig of 73924 bp in length.

FEATURES
Location/Qualifiers
1. 201734
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-693A18"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1. 45183
/note="assembly_fragment"
clone_end:SP6
vector_side:left
45284..46768
/note="assembly_fragment"
46869..47961
/note="assembly_fragment"
48062..49643
/note="assembly_fragment"
49744..51393
/note="assembly_fragment"
51494..74531
/note="assembly_fragment"
74632..127710
/note="assembly_fragment"
127811..201734
/note="assembly_fragment"
clone_end:IT7
vector_side:right"
BASE COUNT 61611 a 39444 c 39572 g 60399 t 708 others
ORIGIN
Query Match 98.5%; Score 646.2; DB 2; Length 201734;
Best Local Similarity 98.8%; Pred. No. 3.8e-146;
Matches 648; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 1 GCCAACAAACAAATGAATAGACCTGGGATGATTTTTGGCCAAAGGCAATTAGAAAA 60
Db 148069 GCCAACAAACAAATGAATAGACCTGGGATGATTTTTGGCCAAAGGCAATTAGAAAA 148010
QY 61 TGATTAGTATCCTTATCAGGAGCAATTCAGACAATGTTGGTGGAGCTTAACACAG 120
Db 148009 TGATTAGTATCCTTATCAGGAGCAATTCAGACAATGTTGGTGGAGCTTAACACAG 147950
QY 121 TGGAGTCAACAGCTGAATCAACGGTGAACAAAGGACAAATAGCCAAATGTACACTTTTAT 180
Db 147949 TGGAGTCAACAGCTGAATCAACGGTGAACAAAGGACAAATAGCCAAATGTACACTTTTAT 147890
QY 181 AAAAACCCACTCCCAAGGACCAGGCACTGGCCCTCTCTCCGGTGCCACAGACATCCACA 240
Db 147889 AAAAACCCACTCCCAAGGACCAGGCACTGGCCCTCTCTCCGGTGCCACAGACATCCACA 147830
QY 241 CAGGCCCAAGATCAGGATTCACAAAGCCAGACAGCAATCGAACGGTTCTGAGTCAFTCG 300
Db 147829 CAGGCCCAAGATCAGGATTCACAAAGCCAGACAGCAATCGAACGGTTCTGAGTCAFTCG 147770
QY 301 CCGGAAGCCTTCCTCAATCAAGCGGAGCTGAAGCATCTCAAGAGGAGCAATAGTCAA 360

```


* 174565	175680:	contig of 1116 bp in length
* 175681	175780:	gap of 100 bp
* 175781	176918:	contig of 1138 bp in length
* 176919	177018:	gap of 100 bp
* 177019	178822:	contig of 1804 bp in length
* 178823	178922:	gap of 100 bp
 Query Match 98.0%; Score 643; DB 2; Length 191793;		
Best Local Similarity 98.5%; Pred.No. 2.3e-145;		
Matches 646;	Conservative 1;	Mismatches 9; Indels 0; Gaps
QY	1	GCCACCAACAATAAGCAATAGACCTGGGATGTATTTTTGGCCCAAGCGAATTAGAAAA 60 Db 28051 GCCAACAAACAAAATGAATAGAACCCTGGGATGTATTTTTGGCCCAAGCGAATTAGAAAA 28110 QY 61 TGATTAGTACTCCTTATFCAGGAGCAATTTACAGAAATGTTTGSGTGGACGTCFAACTACAG 120 Db 28111 TGAATTAGTATCTTATFCAGGAGCAATTTACAGAAATGTTTGSGTGGACGTCFRACTACAG 28170 QY 121 TGGAGTCAAACGTGAATCAAGCGTGAAAAAGCACAAATAGCCAATGTGTACACTTTTTAT 180 Db 28171 TGGAGTCAAACGTGAATCAACGGTGAAAAAGSACAAATAGCCAATGTGTACACTTTTTAT 28230 QY 181 AAAAACACACCTCCAGGACACAGGCATCTGCCCTCTCTCCGGTGCCOCACAGACATCCACA 240 Db 28231 AAAAACACACCTCCAGGACACAGGCATCTGCCCTCTCTCCGGTGCCOCACAGACATCCACA 28290 QY 241 CAGGCCCAAGAAATCAGGATGTGCACAGCCACAGACAAATCCAGCGGTTCTGAGTCACTCTG 300 Db 28291 CAGGCCCAAGAAATCAGGATGTGCACAGCCACAGACAAATPCGAAAGGTTCTGAGTCACTCTG 28350 QY 301 CCGGAAGCCTTGCCCTCAATCAAGCGGACGTGAACATCTCAAAGAGAGGAATAGTCAA 360 Db 28351 CGGGAAGCCTTGCCCTCAATCAAGCGGACGTGAACATCTCAAAGAGAGGAATAGTCAA 28410 QY 361 AGCAGCAGCGCGCGCGCGCGCGCGGCGGACGACGACGACGAGAGGTGGGCGGCTCTG 420 Db 28411 AGCAGCAGCGCGCGCGCGCGCGGCGGACGACGACGACGAGAGGTGGGCGGCTCTG 28470 QY 421 CCAGTATCCGGCGGGGACGACGAGAGGTGCCAGGTTCGCGGAGGAGCCACTCTTCCC 480 Db 28471 CCAGTATCCGGCGGGGACGACGAGAGGTGCCAGGTTCGCGGAGGAGCCACTCTTCCC 28530 QY 481 CTGGAGTGCCTGTAGAGAGGGGAGGAGGAGCCACAGACAGGAATCAGACGCGAGGCAAA 540 Db 28531 CTGGAGTGCCTGTAGAGAGGAGGAGGAGGAGGAGCCACAGACAGGAATCAGACGCGAGGCAAA 28590 QY 541 GCGCGGCAAGGAATCANGAAGATGACGCGGAGCGCGCGCGGAAGAAATCTCGGGGCT 600 Db 28591 GCGCGGCAAGGAATCANGAAGATGACGCGGAGCGCGCGCGGAGAGAGTCCGCGTCT 28650 QY 601 GTGGGGGTTCNCTTGGCACACGCGGGGTCCTCCAAAGCCCCACCGGAGACCCCCGCA 656 Db 28651 GTGGGGGTTCGCGCTTGGCACACGCGGGGTCCAAAGCCCCACCGGAGACCCCCGCA 28706

RESULT 6	AX346386	13202 bp	DNA	linear	PAT 01-FEB-2002
LOCUS	AX346386	Sequence	1457 from Patent WO0200928.		
DEFINITION	AX346386	ACCESSION			
KEYWORDS	AX346386.1	GI:18494272			
SOURCE		synthetic construct.			
ORGANISM		synthetic construct			
		artificial sequences.			
REFERENCE	1				
AUTHORS	Olek, A., Pispembrock, C. and Berlin, K.				
TITLE	Diagnosis of diseases associated with the immune system				
JOURNAL	Parent: WO 0200928-A 1457 03-JAN-2002;				
	Epigenomics AG (DE)				
FEATURES		Location/Qualifiers			
		1. .13202			

```
/organism="synthetic construct"
/db_xref="taxon:32630"
/notes="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      3470 a 187 c 2897 g 6648 t
ORIGIN
Query Match      67.3%; Score 441.6; DB 6; Length 13202;
Best Local Similarity 79.5%; Pred. No. 1.5e-96;
Matches 519; Conservative 1; Mismatches 133; Indels 0; Gaps 0;
OY 4 AACAAACAATGAATGAACCTGGGATGATTTTGGCCAGGCAATTAAGAAATGA 63
Db 4670 AATAAATAAATAAGATTTGGGATGATTTTGGTAAAGTAATTAAGAAATGA 4729
OY 64 TTAGTATCCTTATCAGGAGCAATTCAGAGCAATGTTGGGTGGACGTCACATGCG 123
Db 4730 TTAGTATTTTATAGGAGTAATTTAGAGCAATGTTGGGTGGACGTCACATGCG 4789
OY 124 AGTCAACCTGATCAACCGGTGAATAAGGCAATAGCAATGTGTACATTTTATATA 183
Db 4790 AGTTAAACCTGATTAACCGGTGAATAAGGCAATAGTAATGTGTATTTTATATA 4849
OY 184 AACACCTTCAAGACACGACCTGGCCCTCTCCCGTCCGACACAGATCCACAG 243
Db 4850 AATTATTTTAAAGGATAGTATGTTGTTTTTTTTCGGTCTTTATAGATATTATAG 4909
OY 244 GCCCAAGAAATCAGGATTGCACAGCCAGACCAATCAACGGTTCAGTCAATCTGCG 303
Db 4910 GTTAAAGATTAAGGATTTGTAAGTATAGTATCAACCGTTCAGTCAATCTGCG 4969
OY 304 GAAGCCTTCCCTCAATCAAGCGGAGTGGAAGCATCTACAAAGAGGAATAGTCAAG 363
Db 4970 GAAGTTTCTGTTTAAATAAGCGGACGCTCAAGTATTTTAAAGGAGGAATAGTAAAGT 5029
OY 364 AGCAGCGCGCGCGCGCGGACGACGACGACGACGACGACGACGACGACGACGACG 423
Db 5030 AGTAGCGCGCGCGCGCGGAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 5089
OY 424 GGTACCGCGCGCGGACGACGAGGTGCCAGGTTCGCCGGGAGGACCTTCCCTG 483
Db 5090 GGTATCGCGCGGTTAGTACGAGGTGTTTAGTTTCGCGGAGGTATTTTTCGTTTG 5149
OY 484 GAGTGGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 543
Db 5150 GAGTGGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5209
OY 544 GGCAGCACTANGAATGATCSCGCGGAGGCGCGGCGGAGGAGGAGGAGGAGGAGGAG 603
Db 5210 GGTAGGAATAGGAGATGACGCGCGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAG 5269
OY 604 GGGGTCCCTGGCAGCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 656
Db 5270 GGGGTCCCTGGTATGATGCGGGGTTTTAAGTTTATCGGAGATTCGCGCA 5322

RESULT 7
AP001503
LOCUS      AP001503      186351 bp      DNA      linear      HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-879N20 map 18q22, WORKING
DRAFT SEQUENCE, 23 unordered pieces.
ACCESSION      AP001503
VERSION        AP001503.2 GI:8117361
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Homo sapiens
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 186351)
AUTHORS        Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Iada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE          Homo sapiens 186,351 genomic DNA of 18q22
JOURNAL        Published Only in DataBase (2000)
REFERENCE      2 (bases 1 to 186351)
```

AUTHORS
TITLE
JOURNAL

Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Iada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (17-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsr.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)

COMMENT

On May 30, 2000 this sequence version replaced gi:7288199.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-879N20
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 162162 bases at least Q40
Consensus quality: 175389 bases at least Q30
Consensus quality: 181326 bases at least Q20
Insert size: 184151; sum-of-contigs
Quality coverage: 4.54x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
23 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 45519 contig of 42519 bp in length
42620 59552 contig of 17233 bp in length
59553 74496 contig of 14544 bp in length
74597 87440 contig of 12844 bp in length
87541 99577 contig of 12037 bp in length
99678 110659 contig of 10982 bp in length
115413 115413 contig of 4654 bp in length
110760 124990 contig of 9477 bp in length
115514 135744 contig of 10654 bp in length
125091 145972 contig of 10128 bp in length
135845 153170 contig of 7098 bp in length
146073 156398 contig of 3128 bp in length
153271 161724 contig of 5226 bp in length
156499 165688 contig of 3864 bp in length
161825 165698 contig of 3910 bp in length
165789 173554 contig of 3756 bp in length
169799 175764 contig of 2110 bp in length
173655 177303 contig of 1439 bp in length
175865 179097 contig of 1694 bp in length
177404 181636 contig of 2439 bp in length
179198 183644 contig of 1908 bp in length
181737 184761 contig of 1017 bp in length
183745 184761 contig of 1490 bp in length
184862 186351 contig of 1490 bp in length

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 42519: contig of 42519 bp in length
* 42520 42619: gap of 100 bp
* 42620 59852: contig of 17233 bp in length
* 59853 59952: gap of 100 bp
* 59953 74496: contig of 14544 bp in length
* 74497 74596: gap of 100 bp

	Query Match	57.3%;	Score 376;	DB 2;	Length 186351;
	Best Local Similarity	96.7%;	Pred. No. 1.5e-80;	Indels	Gaps
	Matches 406;	Conservative	0;	Mismatches 10;	
misc_feature	/note="assembly_fragment"				
misc_feature	165789..169498				
misc_feature	/note="assembly_fragment"				
misc_feature	169799..173554				
misc_feature	/note="assembly_fragment"				
misc_feature	173655..175764				
misc_feature	/note="assembly_fragment"				
misc_feature	175865..177303				
misc_feature	/note="assembly_fragment"				
misc_feature	177404..179097				
misc_feature	/note="assembly_fragment"				
misc_feature	179198..181636				
misc_feature	/note="assembly_fragment"				
misc_feature	181737..183644				
misc_feature	/note="assembly_fragment"				
misc_feature	183745..184761				
misc_feature	/note="assembly_fragment"				
misc_feature	184862..186351				
BASE COUNT	56377 a 36459 c 35952 g 55363 t				2200 others
ORIGIN					
QY	1	GCACACAAACAAATCAAAATAGACCTGGGATGTATTTTTTGGCCACAGCAATTAGAAAA	60		
Db	87021	GCACACAAACAAATCAAAATAGACCTGGGATGTATTTTTTGGCCACAGCAATTAGAAAA	87080		
QY	61	TGATTAGTATCTTTATCAGAGCAATTTTCAGAGAATGTTTGGTGGAGCTCTAATCTACAG	120		
Db	87081	TGATTAGTATCTTTATCAGAGCAATTTTCAGAGAATGTTTGGTGGAGCTCTAATCTACAG	87140		
QY	121	TGGAGTCACACGTGAATCAACGGTCGAAAGAGACAAATACCAATGTGTACACTTTTAT	180		
Db	87141	TGGAGTCACACGTGAATCAACGGTCGAAAGAGACAAATACCAATGTGTACACTTTTAT	87200		
QY	181	AAAAACACCCCTCCAGGACACAGGCACCTGGCCCTCTCTCCGGTGGCCACAGACATCCACA	240		
Db	87201	AAAAACACCCCTCCAGGACACAGGCACCTGGCCCTCTCTCCGGTGGCCACAGACATCCACA	87260		
QY	241	CAGGCCCAAGAAATCAGGATTGTCACAAAGCCAGAGCAATCGAACGGTTCTGAGTCACTG	300		
Db	87261	CAGGCCCAAGAAATCAGGATTGTCACAAAGCCAGAGCAATCGAACGGTTCTGAGTCACTG	87320		
QY	301	CGGGAAGCTTGCCCTCAATCAAGCGG ---ACGTGAACATCTCAAGAGGAGATAGT	357		
Db	87321	CGGGAAGCTTGCCCTCAATCAAGCGGCGGACGCTTGATGATCTCAAGAGGAGATAGT	87380		
QY	358	CAAG -CAGCAGCGGGCGGCGCGGCGGACAGACAGCAGCAGAGGTTGGGGGCC	416		
Db	87381	CAAGCCAGCAGCGCGGCTGCGCGGCGGCGGACAGCAGCAGCAGAGGTTGGGGGCC	87440		

```
156499. .161724
/note="assembly_fragment"
161825. .165688
```

```

RESULT 8
AX346387/c
LOCUS
DEFINITION
Sequence 1458 from Patent WO0200928.
ACCESSION
AX346387
VERSION
AX346387.1 GI:18494273
KEYWORDS
synthetic construct.
SOURCE
synthetic construct.
ORGANISM
artificial sequences.
REFERENCE
1
AUTHORS
Olek.A., Piepenbrock,C. and Berlin,K.
TITLE
Diagnosis of diseases associated with the immune system
JOURNAL
Patent: WO 0200928-A 1458 03-JAN-2002;
FEATURES
Epigenomics AG (DE)
Location/Qualifiers
1..13202
source
/organism="synthetic construct"

```

```
/db xref="taxon:32630"
/notes="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3984 a 187 c 2851 g 6180 t
ORIGIN
Query Match 56.1%; Score 367.8; DB 6; Length 13202;
Best Local Similarity 72.3%; Pred. No. 1.2e-78;
Matches 474; Conservative 1; Mismatches 181; Indels 0; Gaps 0;
QY 1 GCCACACAACAAATGAATGAACCTGGGATGTATTTTGGCCGCAAGCAATTAGAAAA 60
Db 8536 GCCACACAACAAATGAATGAACCTGGGATGTATTTTGGCCGCAAGCAATTAGAAAA 8477
QY 61 TGATTAGTATCTTATCAGCAGCAATTCAGAGATGTTGGTGAGCTCTAATCTACAG 120
Db 8476 TGAATATATCTTATCAGCAGCAATTCAGAGATGTTGGTGAGCTCTAATCTACAG 8417
QY 121 TGGAGTCAAGAGTGAATCAACGGTGAAGAAAGGACATAGCAATGTGACACTTTTAT 180
Db 8416 TAAATCAAGAGTGAATCAACGGTGAAGAAAGGACATAGCAATGTGACACTTTTAT 8357
QY 181 AAAAGCACCCTCCAGAGCAGCAGGACTGGCCCTCTCTCCGGTGCCACAGACATCCACA 240
Db 8356 AAAAGCACCCTCCAGAGCAGCAGGACTGGCCCTCTCTCCGGTGCCACAGACATCCACA 8297
QY 241 CAGGCCCAAGAAATCAGGAGTTCGACACGACGAGCAATGAAAGGTTCTGAGTCACTG 300
Db 8296 CAACCCCAAAATCAAAATTTACACACACCAACCAATCGACGATTTCAATCACTA 8237
QY 301 CCGAAGACCTTCCCTCAATCAAGCGGACGTGAAGCAATCAAAAGGAGGAATAGTCAA 360
Db 8236 CCGAAGACCTTCCCTCAATCAAGCGGACGTGAAGCAATCAAAAGGAGGAATAGTCAA 8177
QY 361 AGCAGACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGT 420
Db 8176 AACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 8117
QY 421 CCAGGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 480
Db 8116 CCAATACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 8057
QY 481 CTGAGTGGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540
Db 8056 CTAATATAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7997
QY 541 GCGGCGCAGGAACTAAGAGATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGT 600
Db 7996 AACGACAAACAACTAAGAGATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGAA 7937
QY 601 GTGGGGTCCNCCCTGGCACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGGA 656
Db 7936 ATAAAAATCGCCTTAACCAACCAAGCAATCCCAAAACCCCAACCCGCGGAGCCG 7881
RESULT 9
AR005195 1425 bp DNA linear PAT 04-DEC-1998
LOCUS
DEFINITION Sequence 1 from patent US 5747645.
ACCESSION AR005195
VERSION AR005195.1 GI:3966074
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Sprecher, C.A.
TITLE Cytoplasmic antiproteinase-2 and cytoplasmic antiproteinase-3 and coding sequences
JOURNAL Patent: US 5747645-A 1 05-MAY-1998;
FEATURES Location/Qualifiers
source
BASE COUNT 437 a 301 c 357 g 330 t
ORIGIN
Query Match 11.2%; Score 73.4; DB 6; Length 1425;
Best Local Similarity 98.7%; Pred. No. 3.2e-07;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 333 GAAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCG 392
Db 7 GGAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCGG 66
QY 393 CAGCAGCAGCAGGAG 407
Db 67 CAGCAGCAGCAGGAG 81
RESULT 10
I81228 1425 bp DNA linear PAT 10-JUN-1998
LOCUS
DEFINITION Sequence 1 from patent US 5710026.
ACCESSION I81228
VERSION I81228.1 GI:3209518
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Sprecher, C.A.
TITLE Cytoplasmic antiproteinase-2 and cytoplasmic antiproteinase-3 and coding sequences
JOURNAL Patent: US 5710026-A 1 20-JAN-1998;
FEATURES Location/Qualifiers
source
BASE COUNT 437 a 301 c 357 g 330 t
ORIGIN
Query Match 11.2%; Score 73.4; DB 6; Length 1425;
Best Local Similarity 98.7%; Pred. No. 3.2e-07;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 333 GAAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGG 392
Db 7 GGAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCGG 66
QY 393 CAGCAGCAGCAGGAG 407
Db 67 CAGCAGCAGCAGGAG 81
RESULT 11
I82208 1425 bp DNA linear PAT 10-JUN-1998
LOCUS
DEFINITION Sequence 1 from patent US 5712117.
ACCESSION I82208
VERSION I82208.1 GI:3210505
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Sprecher, C.A.
TITLE Cytoplasmic antiproteinase-2 and coding sequences
JOURNAL Patent: US 5712117-A 1 27-JAN-1998;
FEATURES Location/Qualifiers
source
BASE COUNT 437 a 301 c 357 g 330 t
ORIGIN
Query Match 11.2%; Score 73.4; DB 6; Length 1425;
Best Local Similarity 98.7%; Pred. No. 3.2e-07;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 333 GAAGCATCTACAAGAGGAGTAATAGTCAAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGAG 392
 Db 7 GGAGCATCTACAAGAGGAGTAATAGTCAAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGAG 66

QY 393 CAGCAGCAGCAGGAG 407
 Db 67 CAGCAGCAGCAGGAG 81

RESULT 12
 HUMCAP2A 1325 bp mRNA linear PRI 18-JAN-1996
 LOCUS Homo sapiens cytoplasmic antiproteinase 2 (CAP2) mRNA, complete cds.
 DEFINITION L40377
 VERSION L40377.1 GI:1160926
 KEYWORDS cytoplasmic antiproteinase; cytoplasmic antiproteinase 2; ovalbumin serine proteinase inhibitor 8; serine proteinase inhibitor.
 SOURCE Homo sapiens placenta cDNA to mRNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1325)
 AUTHORS Sprecher,C.A., Morgenstern,K.A., Mathewes,S., Dahlen,J.R., Schrader,S.K., Foster,D.C. and Kiesel,W.
 TITLE Molecular cloning, expression, and partial characterization of two novel members of the ovalbumin family of serine proteinase inhibitors
 JOURNAL J. Biol. Chem. 270 (50), 29854-29861 (1995)
 MEDLINE 96102039
 PUBMED 8530382
 FEATURES
 source Location/Qualifiers
 1..1325
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="placenta"
 /note="(vector lambda gt11)"
 1..1325
 /gene="CAP2"
 1..1325
 /gene="CAP2"
 1..83
 /gene="CAP2"
 84..1208
 /gene="CAP2"
 /codon_start=1
 /product="cytoplasmic antiproteinase 2"
 /protein_id="AAC41939.1"
 /db_xref="GI:1160927"
 /translation="MDLCEANGTFATSLFKILGEDNSRNVFFSPMSISSALAWFM GAGSTAAQMSQALCIYKDGIHRGFSQSLSEVNRGTQYLLRANLFGKTCDFLP DKKEYCKQFQALBELSFABDETECKRHNDVAKTEGKISEVLDAVTYDPTKLV LVNAIFYKGNWEPDRKYTRGMLEKTNBKKTVOMKFEAKFMGYADVHTQVLEL PYVEELSWILLPDNDLAVKEALITYEKFKAWTNSEKLKSKYQVFLPKLKEES YDLFFPLRLGLMDAPDAKADFSGMSTERNVPLSKVHAKCFEYVNEEGTEAANAATV VNRSRMRPEFPCADHPFLFFIRRHKNCILFCGRFSSP"
 1209..1325
 /gene="CAP2"
 1307..1312
 /partial
 /gene="CAP2"
 3'UTR
 polyA_signal
 BASE COUNT 378 a 286 c 342 g 319 t
 ORIGIN

Query Match 11.18; Score 73; DB 9; Length 1325;
 Best Local Similarity 100.0%; Pred. No. 4e-07;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 AGCATCTACAAGAGGAGTAATAGTCAAGCAGCAGCGCGCGCGCGCGCGCGCGCGAGCA 394
 Db 1 AGCATCTACAAGAGGAGTAATAGTCAAGCAGCAGCGCGCGCGCGCGCGCGAGCA 60

QY 395 GCAGCAGCAGGAG 407

Db 61 GCAGCAGCAGGAG 73

RESULT 13
 I66494/c 7218 bp DNA linear PAT 28-DEC-1997
 LOCUS Sequence 14 from patent US 5670367.
 DEFINITION I66494
 ACCESSION I66494
 VERSION I66494.1 GI:2724471
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 7218)
 AUTHORS Dörner,F., Scheifflinger,F. and Falkner,F.Günter.
 TITLE Recombinant fowlpox virus
 JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
 FEATURES Location/Qualifiers
 1..7218
 /organism="unknown"
 BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
 ORIGIN

Query Match 11.0%; Score 72; DB 6; Length 7218;
 Best Local Similarity 3.8%; Pred. No. 7.5e-07;
 Matches 15; Conservative 234; Mismatches 142; Indels 0; Gaps 0;

QY 248 AAGAATCAGGATTCACAAGCCAGCAGCAATCGAAGCTTCTGAGTCATCTCCCGAAG 307
 Db 1425 RRR 1366

QY 308 CCTTGCCCTCAATCAAGCGGACGTGAAGCATCTACAAGAGGAGGATAGTCAAGCAGCA 367
 Db 1365 RRR 1306

QY 368 CGCGCGCGCGCGCGCGCGCAGCAGCAGCAGCAGGAGTGGGGCCCTCCAGGTA 427
 Db 1305 RRR 1246

QY 428 CCGCGCGCGCAGCAGGAGTCCCAGGTTCGCCGCGAGGCCACCTCTCCTCGAGT 487
 Db 1245 RRR 1186

QY 488 GCGTGAGAGGAGGAGGAGGAGCAGCAGCAGGATCAGAGCGAGCAAGCGCGGC 547
 Db 1185 RRR 1126

QY 548 AGGAATAGAGATGACGCGGCGGCGCGCGGAGAAAGAAANTCTCGGGCTGTGGGG 607
 Db 1125 RRR 1066

QY 608 TCNCCCTGGCAGCAGCGGGTCCCAAGCC 638
 Db 1065 TCGCAAGCTCCCTCGACCTCGAGCAAGCTC 1035

RESULT 14
 AC103021 113684 bp DNA linear HTG 13-JUL-2002
 LOCUS Rattus norvegicus clone CH230-197M15, *** SEQUENCING IN PROGRESS
 DEFINITION *** 65 unordered pieces.
 ACCESSION AC103021
 VERSION AC103021.3 GI:21730130
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 113684)
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1049: contig of 1049 bp in length
 1050 1149: gap of unknown length
 1150 2299: contig of 1150 bp in length
 2300 2399: gap of unknown length
 2400 4013: contig of 1614 bp in length
 4014 4113: gap of unknown length
 4114 5129: contig of 1016 bp in length
 5130 5229: gap of unknown length
 5230 6293: contig of 1064 bp in length
 6294 6393: gap of unknown length
 6394 7511: contig of 1118 bp in length
 7512 7612: gap of unknown length
 7613 8743: contig of 1132 bp in length
 8744 8843: gap of unknown length
 8844 10007: contig of 1164 bp in length
 10008 10107: gap of unknown length
 10108 11279: contig of 1172 bp in length
 11280 11379: gap of unknown length
 11380 12440: contig of 1061 bp in length
 12441 12540: gap of unknown length
 12541 13991: contig of 1351 bp in length
 13992 13991: gap of unknown length
 13992 15136: contig of 1145 bp in length
 15137 15236: gap of unknown length
 15237 16497: contig of 1261 bp in length
 16498 16597: gap of unknown length
 16598 17707: contig of 1110 bp in length
 17708 17807: gap of unknown length
 17808 19233: contig of 1426 bp in length
 19234 19333: gap of unknown length
 19334 20707: contig of 1374 bp in length
 20708 21907: contig of 1100 bp in length
 21908 22007: gap of unknown length
 22008 23207: contig of 1200 bp in length
 23208 23307: gap of unknown length
 23308 24909: contig of 1602 bp in length
 24910 25009: gap of unknown length
 25010 26420: contig of 1411 bp in length
 26421 26520: gap of unknown length
 26521 28020: contig of 1500 bp in length
 28021 28120: gap of unknown length
 28121 29722: contig of 1602 bp in length
 29723 29822: gap of unknown length
 29823 31076: contig of 1254 bp in length
 31077 31176: gap of unknown length
 31177 32290: contig of 1114 bp in length
 32291 32390: gap of unknown length
 32391 33610: contig of 1220 bp in length
 33611 33710: gap of unknown length
 33711 34987: contig of 1277 bp in length
 34988 35087: gap of unknown length
 35088 36268: contig of 1181 bp in length
 36269 36368: gap of unknown length
 36369 38071: contig of 1703 bp in length
 38072 38171: gap of unknown length
 38172 39859: contig of 1688 bp in length
 39860 39959: gap of unknown length
 39960 41056: contig of 1097 bp in length
 41057 41156: gap of unknown length
 41157 42335: contig of 1179 bp in length
 42336 42435: gap of unknown length
 42436 44405: contig of 1970 bp in length
 44406 44505: gap of unknown length
 44506 46527: contig of 2022 bp in length
 46528 46627: gap of unknown length
 46628 48152: contig of 1525 bp in length

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
 Burch, P., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Caron, T., Carter, K., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Dean, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, R.,
 Honsi, F., Howard, S., Huber, J., Hulky, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Z., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegad, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawlin, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, J., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sison, I.,
 Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GIYQ
 Center clone name: CH230-197M15
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 46493 bases at least Q40
 Consensus quality: 49763 bases at least Q30
 Consensus quality: 51555 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 65 contigs. The true order of the pieces

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi, Muridae; Murinae; Mus.
1 (bases 1 to 65711)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-133N21
Unpublished
2 (bases 1 to 65711)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouckhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gort,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
McLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K.K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodor,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 65711)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouckhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murry,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 13, 2002 this sequence version replaced gi:20128326.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>

```

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
*
-----
Center project name: 122422
Center clone name: 133_N_21

```


Fri Jan 10 08:43:24 2003

us-09-581-500b-12.rge

Page 15

Search completed: January 10, 2003, 00:16:31
Job time : 2751.88 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 19:55:35 ; Search time 311.247 Seconds
(without alignments)
4746.432 Million cell updates/sec

Title: US-09-581-500B-12
Perfect score: 656
Sequence: 1 gccacacacacaaatgaat.....ccacccgcgagaccgcgga 656

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	652.6	99.5	656	AA88553	Human chromosome 1
2	441.6	67.3	13202	ABL33484	Human immune syste
3	367.8	56.1	13202	ABL33485	Human immune syste
4	100	15.2	186	AAC13994	Human secreted pro
5	73.4	11.2	1425	17 AAT35220	Cytoplasmic antipr
6	73.4	11.2	1425	21 AAZ39749	Human cytoplasmic
7	73	11.1	1325	2 AAK83894	Human cDNA differe
8	50.4	7.7	10732	21 AAL10594	Gene encoding a su
9	50.2	7.7	795	19 AAV55830	FLGA insert stabil

10	49.2	7.5	1925	20 AAX90924	Estein Barr Virus
11	47.2	7.2	3103	23 AAS85091	DNA encoding novel
12	46.8	7.1	5222	21 AAA55964	Human G713 5'-regu
13	46.8	7.1	5566	21 AAA55967	Human G713 5'-regu
14	46.4	7.1	286	16 AAZ95180	Simple tandem repe
15	46.4	7.1	1017	20 AAZ17058	Human gene expres
16	46	7.0	916	22 AAF80065	Amino terminus of
17	46	7.0	3563	20 AAZ23429	Human SCA6 DNA fra
18	46	7.0	3596	19 AAV61588	Alpha-1A calcium c
19	46	7.0	3632	19 AAV61586	Alpha-1A calcium c
20	45.6	7.0	599	22 ABA60694	Human foetal liver
21	45.6	7.0	599	22 ABA63254	Human foetal liver
22	45.6	7.0	599	22 ABA28778	Probe #7244 for ge
23	45.6	7.0	599	22 ABA30484	Probe #9250 for ge
24	45.6	7.0	599	22 AAK08976	Human brain expres
25	45.6	7.0	599	22 AAK11722	Human brain expres
26	45.6	7.0	599	22 AAK34867	Human bone marrow
27	45.6	7.0	599	22 AAK37465	Human bone marrow
28	45.6	7.0	599	22 AAI16860	Probe #6793 for ge
29	45.6	7.0	599	22 AAI18270	Probe #9203 for ge
30	45.6	7.0	599	22 AAI40583	Probe #9269 used t
31	45.6	7.0	599	22 AAI43326	Probe #12012 used
32	45.6	7.0	599	24 ABS09398	Human genome-deriv
33	45.6	7.0	599	24 ABS11444	Human genome-deriv
34	45.6	7.0	840	24 ABQ35494	Oligonucleotide fo
35	45.6	7.0	840	24 ABQ35495	Oligonucleotide fo
36	45.6	7.0	855	22 AAH03166	Human cDNA clone (
37	45.6	7.0	1881	22 AAH13633	Human cDNA sequenc
38	45.6	7.0	2027	24 ABN59664	Novel human coding
39	45.6	7.0	6545	23 ABL14373	Drosophila melanog
40	45.6	7.0	10438	23 ABL14372	Drosophila melanog
41	45.4	6.9	799	19 AAV55831	Nucleotide sequenc
42	45.4	6.9	1926	21 AAA50254	Epstein Barr virus
43	45.4	6.9	1926	22 AAF82902	EBV tethering prot
44	45.4	6.9	2493	23 AAS88122	DNA encoding novel
45	45.4	6.9	2580	21 AAT75454	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AX88553
ID AAX88553 standard; DNA: 656 BP.
XX AAX88553;
XX
DT 10-SEP-1999 (first entry)
DE Human chromosome 18q YAC clone nucleotide sequence #12.
XX
KW Human chromosome 18q; mood disorder; polymorphic marker; detection;
KW identification; trinucleotide repeat expansion; schizophrenia;
KW anxiety disorder; adjustment disorder; personality disorder;
KW nucleotide triplet repeat; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO9932643-A2.
XX
PD 01-JUL-1999.
XX
PF 17-DEC-1998; 98WO-EP08543.
XX
PR 18-DEC-1997; 97GB-0026804.
PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX
PI Del-Favero J, Raeynaekers P, Van Broeckhoven C;
XX
DR WPI; 1999-418934/35.
XX

XX PI Sprecher CA, Foster DC, Jaspers SR;
 XX DR WPI; 2000-052146/05.
 XX P-PSDB; AAY55840.
 XX Method for treating disease or symptoms of a disease mediated by a
 PT caspase -
 PT caspase -
 XX Example 1; Page 55-56; 65pp; English.
 XX The invention provides a method for treating a disease mediated by a
 CC caspase in an individual. The method comprises: administering a
 CC composition comprising a gene coding for an intracellular mammalian
 CC serpin in an amount sufficient to inhibit activity of the caspase upon
 CC transient expression of the gene in a target tissue affected by the
 CC disease, where the disease or the symptoms are treated. The method can be
 CC used for decreasing inflammation, for modulating apoptosis, for treating
 CC a lung disease, and for treating a neurodegenerative disease. The
 CC inflammation and apoptosis that can be treated are particularly in heart
 CC or liver tissue. It can be used for treating Alzheimer's disease,
 CC Parkinson's disease, amyotrophic lateral sclerosis, and acute injury
 CC such as hypoxia-ischemia or trauma. The present sequence represents a
 CC cDNA encoding the human cytoplasmic antiproteinase-2 protein (CAP-2), an
 CC intracellular serpin protein.
 XX Sequence 1425 BP; 437 A; 301 C; 357 G; 330 T; 0 other;
 SQ

Query Match 11.2%; Score 73.4; DB 21; Length 1425;
 Best Local Similarity 98.7%; Pred. No. 9.5e-09;
 Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 333 GAAGCATCTCAAAAGGAGGATAGTCAAGCAGCAGCGCGCGCGCGCGCGCAGCAG 392
 Db 7 GGAGCATCTCAAAAGGAGGATAGTCAAGCAGCAGCGCGCGCGCGCGCGCAGCAG 66
 QY 393 CAGCAGCAGCAGGAG 407
 Db 67 CAGCAGCAGCAGGAG 81
 RESULT 7
 ABE83894
 ID ABE83894 standard; cDNA; 1325 BP.
 AC ABE83894;
 XX 14-AUG-2002 (first entry)
 XX Human cDNA differentially expressed in granulocytic cells #465.
 DE Human; ss; granulocytic cell; DNA chip; bacterial infection;
 XX viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.
 XX Homo sapiens.
 OS
 XX WO200228999-A2.
 PN 11-APR-2002.
 PD 03-OCT-2001; 2001WO-US30821.
 PF 03-OCT-2000; 2000US-237189P.
 XX (GENE-) GENE LOGIC INC.
 PR Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX

XX WPI; 2002-435328/46.
 DR Detecting granulocyte activation by detecting differential expression
 XX of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 XX Claim 1; SEQ ID No 465; 114pp; English.
 PS The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) Gs by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating Gs; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 1325 BP; 378 A; 286 C; 342 G; 319 T; 0 other;
 Query Match 11.1%; Score 73; DB 24; Length 1325;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 335 AGCATCTCAAAAGGAGGATAGTCAAGCAGCAGCGCGCGCGCGCGCGCAGCAGCA 394
 Db 1 AGCATCTCAAAAGGAGGATAGTCAAGCAGCAGCGCGCGCGCGCGCGCAGCAGCA 60
 QY 395 GCAGCAGCAGGAG 407
 Db 61 GCAGCAGCAGGAG 73
 RESULT 8
 AAA10594
 ID AAA10594 standard; DNA; 10732 BP.
 XX AAA10594;
 AC AAA10594;
 XX 29-JUN-2000 (first entry)
 DT Gene encoding a subunit of cellulose synthase.
 DE Cellulose synthase; cellulose production; increase yield; ds.
 XX

XX Vigna angularis.
OS JP2000060568-A.
PN 29-FEB-2000.
XX 26-AUG-1998; 98JP-0239998.
XX 26-AUG-1998; 98JP-0239998.
XX (MIZUO) MIZUNO K.
PA (OJIP) OJI PAPER CO.
XX WPI: 2000-342371/30.
DR P-PSDB; AAY85179.
XX
XX A gene encoding a cellulose synthetic equipment - for the improvement
PT in the amount of cellulose synthesised in a plant body
PT
XX
XX Claim 2; Page 14-21; 32pp; Japanese.
XX
XX This sequence represents a gene encoding a subunit of the cellulose
CC synthase complex of Vigna angularis. The invention relates to subunits of
CC cellulose synthetic equipment, that can be used to increase the amount of
CC cellulose synthesised by a plant. The proteins and genes encoding them
CC can also be used to improve the properties of the cellulose being
CC produced by a plant.
XX
XX Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;
SQ

Query Match 7.7%; Score 50.4; DB 21; Length 10732;
Best Local Similarity 21.88; Pred. No. 0.012;
Matches 111; Conservative 132; Mismatches 244; Indels 2; Gaps 1;

QY 146 AAAAAAGGACATCAATGACATGTTTATATAAAACCCCTCCAGGACGAGC 205
DB 8807 AATAAGGAAATTTATTTGTTTTCATATTTTATGAAAAAAGGAGGSRAGG 8866
QY 206 ACTGGCCCTCTCGGTCGCCACACATCCACAGCCCAAGATCAGGATGCA 265
DB 8867 ASTYRGNTSTHRTFRAGYSYSSRGYSTHRAAASASNTHRSTRYSRGSNG 8926
QY 266 CAGCAGACGACATCAAGGCTTCTGATCATCTCGCGAAGCCTTCCCTCAATCAAGG 325
DB 8927 GVAGNAVAASASYSASGTSASYSARGASNSYSASRSASRASCYSASNAYSAS--S 8984
QY 326 CGGACGTGAAGCATCTACAAGGAGGAGTAAAGTCAAGACGAGCGCGCGCGCGCGG 385
DB 8985 NGVAYSNGNTHRYSVAAGGAAATTAAGTYSYSAAGAGVASRMTYSGNASASGAGNG 9044
QY 386 GCAGCAGCAGCAGCAGGAGGAGTGGGGCTTCCAGGTACCGGCGGCGGCGGCGG 445
DB 9045 NARGVATYRGAGVAAAASASGTAAYSGCYSMTNGNAGRHVAARGGGGNGNGHSA 9104
QY 446 GAGTGCCAGCTCCCGGAGGCGACCTCTTCCCTGAGTGGCTGAGAGAGGGAAGG 505
DB 9105 SAAVAMTYSASRYSCHYSARGMTYVAGGNSRGTHRSYRSARGAAYSAGVAASNS 9164
QY 506 GAGGAGGCGGAGCAGGAATCAGAGCGAGGCAAGCGGCGGCGGCGGCGGAGTAC 565
DB 9165 RASASNYSHAAAGGASNGNASARGAGGNTHRGNAAGGASHGYAAMTSRAGGSRM 9224
QY 566 SGCGGAGGCGGCGGGAAGAAATCTCGGGCTGTGGGGCTGCGGCGGCGGCGGCGG 625
DB 9225 TGYASASNTSRTHRTYRGVAGVAGYSGGARGASNGGARGGHSRARGHTRARG 9284
QY 626 GGGTCCCAAGCCCGGAGACCCGCG 654
DB 9285 VASRHSYSGNGSRTHRTYSYSAAYSGCG 9313
RESULT 9

AAV55830
ID AAV55830 standard; DNA; 795 BP.
XX
AC AAV55830;
XX
DT 18-NOV-1998 (first entry)
XX
DE FLGA insert stabilising polypeptide encoding DNA.
XX
XX Fusion protein; stabilising polypeptide; proteolytic degradation;
XX resistance; half-life; autoimmune disease; inflammation; nitro drug;
XX IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
XX nitroreductase protein; enzyme therapy; prodrug therapy; protease;
XX cancer; pathological condition; ss.
OS Epstein-barr virus.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..788
FT /*tag= a
FT /product= "stabilising polypeptide"
XX
XX W09822577-AL.
XX
XX 28-MAY-1998.
XX
XX 17-NOV-1997; 97WO-IB01508.
XX
XX 25-JUN-1997; 97US-0048945.
XX
XX 15-NOV-1996; 96US-0030986.
XX
XX (MASU/) MASUCCI M G.
XX
XX Masucci MG;
XX
XX WPI: 1998-312463/27.
DR P-PSDB; AAW79128.
XX
XX New fusion proteins resistant to proteolytic degradation -
PT comprising a core protein with a stabilising polypeptide comprising
PT a peptide sequence containing glycine repeats
XX
PS Disclosure; Fig 3; 120pp; English.
XX
XX This DNA encodes a stabilising polypeptide and is the FLGA insert of the
CC invention. The invention provides a method for increasing the resistance
CC of a core protein to proteolytic degradation that comprises linking or
CC inserting onto or into the core protein a stabilising polypeptide of
CC formula [(Gly)X(Gly)Y(Gly)Z]_n where Gly, Glyb, Glyc are 1-6
CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC Phe, pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC be identical from n repeat to n repeat. Alternatively a nucleic acid
CC encoding the stabilising polypeptide can be linked onto or inserted into
CC a nucleic acid encoding a core protein. The fusion proteins of the
CC invention are more resistant to degradation by proteases and, thus, have
CC a longer half-life than the unfused core protein. The products can be
CC used for treating autoimmune diseases, cancer and inflammation. In
CC particular, the core protein may be an IkappaB regulator protein for the
CC treatment of inflammatory bowel disease, or a nitroreductase protein
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC or other pathological conditions. The fusion proteins can also be used in
CC diagnostic methods such as in vivo imaging.
XX
SQ Sequence 795 BP; 200 A; 104 C; 478 G; 13 T; 0 other;
XX
XX Query Match 7.7%; Score 50.2; DB 19; Length 795;
XX Best Local Similarity 48.2%; Pred. No. 0.0057;
XX Matches 136; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 346 AGGAGGATAGTCATCAAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 405
DB 192 AGGAGGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 251

QY 406 AGGTGGGGCTCTCCAGTACCGGGGGGCGAGGACGACGAGGTCCAGTTCCTCCGCG 465
AAAX90924
ID AAX90924 standard; DNA; 1925 BP.
XX
AC AAX90924;
XX
DT 17-JAN-2000 (first entry)
XX
DE Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.
XX
KW Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid pCMVEBNA; EBNA 1;
KW episome; transfection; origin of replication; EBV orp; receptor;
KW eucaryotic host cell; recombinant cell line; ion channel; gene therapy;
KW multiple gene expression; transporter protein; transcription factor;
KW adhesion molecule; antisense therapy; gene amplification;
KW cell immortalisation; ds.
XX
OS Epstein-barr virus.
XX
FH Key Location/Qualifiers
FT 1..1925
CDS
FT FT
FT /*tag= a
FT /product= "EBNA 1"
FT /transl_except= (pos:799..800, aa:Gly)
FT /note= "The sequence is described throughout the
FT specification as being 1926 nucleotides long, but a
FT sequence of only 1925 bp has been given in figure 2"
PN WO9947647-A1.
XX
XX 23-SEP-1999.
XX
XX 12-FEB-1999; 99WO-US03307.
XX
XX 18-MAR-1998; 98US-0040961.
XX
XX 06-AUG-1998; 98US-0130114.
XX
XX (PHAR-) PHARMACORPEIA INC.
XX
XX Danaaj BB, Horlick RA, Robbins AK;
PI
XX WPI; 1999-610610/52.
DR
XX P-PSDB; AAY28843.
XX
XX New method for expressing genes from recombinant eukaryotic cells,
PT useful for gene therapy -
PT
XX Claim 24; Fig 2; 86pp; English.
PS
XX The present sequence is a DNA encoding Epstein Barr Virus Nuclear
CC Antigen 1 (EBNA 1), which is obtained from commercially available
CC plasmid pCMVEBNA. EBNA 1 protein is used to stably maintain episomes
CC containing EBV origin of replication (oriP) and a gene encoding
CC protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
CC protein are transfected with these episomes to produce recombinant
CC cell lines expressing multiple genes of interest. This provides a

CC rapid and reliable method of stably expressing multiple genes in
CC transfected cells. The episomes are useful in the transfection of genes
CC encoding receptors, transporter proteins, ion channels, adhesion
CC molecules and transcription factors. The episomes carrying desired genes
CC can also be used to transfect cells in gene therapy, antisense therapy,
CC for gene amplification, cell immortalisation, etc.
XX
SQ Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;
Query Match 7.5%; Score 49.2; DB 20; Length 1925;
Best Local Similarity 49.8%; Pred. No. 0.014;
Matches 120; Conservative 1; Mismatches 120; Indels 0; Gaps 0;
QY 346 AGGAGGAATAGTCAAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 405
DB 705 AGGAGCAGGAGGGCGCAGGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 764
QY 406 AGGTGGGGCTCTCCAGTACCGGGGGGCGAGGACGACGAGGTCCAGTTCCTCCGCG 465
DB 765 AGGAGGCGCAGGAGGGCGCAGGACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 824
QY 466 GAGGCCACCTCTTCCCTGGAGTGGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 525
DB 825 GAGGCGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 884
QY 526 TCAGAGCGAGGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAA 585
DB 885 GGAGCAGGAGGGCGCAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGCA 944
QY 586 G 586
DB 945 G 945
RESULT 11
AAS85091/C
ID AAS85091 standard; cDNA; 3103 BP.
XX
AC AAS85091;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #20895.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR
XX P-PSDB; ABG20904.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity -
PS Claim 1; SEQ ID No 20895; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and

Search completed: January 9, 2003, 21:33:45
Job time : 327.247 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 21:50:31 ; Search time 55.2338 Seconds
(without alignments)
3642.332 Million cell updates/sec

Title: US-09-581-500B-12
Perfect score: 656
Sequence: 1 gccacaacaaatgaat.....cccacgcgagaccgcgca 656

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73.4	11.2	1425	1 US-08-464-148-1	Sequence 1, Appli
2	73.4	11.2	1425	1 US-08-385-500-1	Sequence 1, Appli
3	73.4	11.2	1425	1 US-08-846-784-1	Sequence 1, Appli
c 4	72	11.0	7218	1 US-08-232-463-14	Sequence 14, Appli
c 5	49.4	7.5	1104	4 US-09-009-816-1	Sequence 1, Appli
6	46.4	7.1	286	2 US-08-332-766A-6	Sequence 6, Appli
7	46	7.0	3563	4 US-09-041-886-20	Sequence 20, Appli
8	46	7.0	3596	2 US-08-779-801-5	Sequence 5, Appli
9	46	7.0	3596	2 US-09-238-441-5	Sequence 5, Appli
10	46	7.0	3632	2 US-08-779-801-3	Sequence 3, Appli
11	46	7.0	3632	4 US-09-238-441-3	Sequence 3, Appli
12	45.8	7.0	325	2 US-08-531-927B-3	Sequence 3, Appli
13	45.4	6.9	1926	4 US-09-249-585A-2	Sequence 2, Appli
14	45.4	6.9	2580	3 US-09-050-863-2	Sequence 2, Appli
15	45.4	6.9	2580	4 US-09-359-081-2	Sequence 2, Appli
c 16	45.4	6.9	5452	2 US-09-130-114-1	Sequence 1, Appli
17	45.4	6.9	9600	4 US-08-910-647-1	Sequence 1, Appli
18	45.4	6.9	9600	4 US-09-620-925-1	Sequence 1, Appli
19	45.4	6.9	10596	1 US-07-884-811-15	Sequence 15, Appli
20	45.4	6.9	10596	1 US-07-885-971-15	Sequence 15, Appli
21	45.4	6.9	10596	1 US-08-087-783A-15	Sequence 15, Appli
22	45.4	6.9	10596	1 US-08-194-088B-15	Sequence 15, Appli
23	45.4	6.9	10596	2 US-08-194-087-15	Sequence 15, Appli
24	45.4	6.9	10596	5 PCT-US93-04648-15	Sequence 15, Appli
25	44.6	6.8	1028	4 US-08-118-200-1	Sequence 1, Appli
26	44.6	6.8	1028	4 US-08-458-745-1	Sequence 1, Appli
27	44.4	6.8	4488	4 US-08-406-030A-3	Sequence 3, Appli

28	44.2	6.7	2056	4 US-09-334-601-12	Sequence 12, Appli
29	43.6	6.6	4362	2 US-08-455-073A-1	Sequence 1, Appli
30	43.6	6.6	7808	2 US-08-149-097D-22	Sequence 22, Appli
31	43.6	6.6	7808	3 US-08-949-386-22	Sequence 22, Appli
32	43.6	6.6	7808	3 US-08-450-562-22	Sequence 22, Appli
33	43.6	6.6	7808	4 US-08-984-709A-22	Sequence 22, Appli
34	43.6	6.6	7808	4 US-08-450-272-22	Sequence 22, Appli
c 35	43	6.6	670	4 US-09-009-816-3	Sequence 3, Appli
36	42.6	6.5	300	4 US-09-135-994-3	Sequence 1, Appli
37	42.2	6.4	477	4 US-09-135-994-1	Sequence 1, Appli
38	42.2	6.4	2115	2 US-08-474-379C-60	Sequence 60, Appli
39	42.2	6.4	2115	3 US-09-146-249A-60	Sequence 60, Appli
40	42.2	6.4	2115	3 US-08-206-188B-60	Sequence 60, Appli
41	42.2	6.4	2770	4 US-09-008-697A-13	Sequence 13, Appli
c 42	42.2	6.4	4897	6 5196516-7	Patent No. 5196516
43	42	6.4	9972	3 US-08-836-022A-3	Sequence 3, Appli
c 44	42	6.4	9972	4 US-09-427-048A-3	Sequence 3, Appli
c 45	42	6.4	53526	3 US-08-658-136-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-464-148-1
; Sequence 1, Application US/08464148
; Patent No. 5710026
; GENERAL INFORMATION:
; APPLICANT: Sprecher, Cindy A.
; TITLE OF INVENTION: CYTOPLASMIC ANTIPROTEINASE-2 AND CODING SEQUENCES
; TITLE OF INVENTION: CYTOPLASMIC ANTIPROTEINASE-3 AND CODING SEQUENCES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourile and Crew
; STREET: Stewart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,148
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/385,500
; FILING DATE: 08-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1425 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 92..1213
; OTHER INFORMATION: /product= "CYTOPLASMIC
; OTHER INFORMATION: ANTIPROTEINASE-2 PROTEIN"
; US-08-464-148-1

Query Match 11.2%; Score 73.4; DB 1; Length 1425;

RESULT 3
US-08-846-784-1

CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500


```

1  RESULT 7
2  US-09-041-886-20
3  ; Sequence 20, Application US/09041886
4  ; Patent No. 6235872
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Bredesen, Dale E.
7  ; APPLICANT: Razibadeh, Sharroz
8  ; TITLE OF INVENTION: Proapoptotic
9  ; TITLE OF INVENTION: Polypeptides
10 ; NUMBER OF SEQUENCES: 72
11 ; CORRESPONDENCE ADDRESS:
12 ; ADDRESSEE: Campbell & Flores L
13 ; STREET: 4370 La Jolla Village
14 ; CITY: San Diego
15 ; STATE: California
16 ; COUNTRY: United States
17 ; ZIP: 92122
18 ; COMPUTER READABLE FORM:
19 ; MEDIUM TYPE: Floppy disk
20 ; COMPUTER: IBM PC compatible
21 ; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

1  RESULT 8
2  US-08-779-801-5
3  ; Sequence 5, Application US/08779801
4  ; Patent NO. 5853995
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Lee, Cheng-Chi
7  ; TITLE OF INVENTION: Large Scale Genotyping of
8  ; TITLE OF INVENTION: Diseases and a Diagnostic
9  ; NUMBER OF SEQUENCES: 5
10 ; CORRESPONDENCE ADDRESS:
11 ; ADDRESSEE: Benjamin Aaron Adler, Ph.D., J.D.
12 ; STREET: 8011 Candle Lane
13 ; CITY: Houston
14 ; STATE: Texas
15 ; COUNTRY: USA
16 ; ZIP: 77071
17 ; COMPUTER READABLE FORM:
18 ; MEDIUM TYPE: Floppy disk
19 ; COMPUTER: Apple Macintosh
20 ; OPERATING SYSTEM: Macintosh
21 ; SOFTWARE: Microsoft Word for Macintosh
22 ; CURRENT APPLICATION DATA:
23 ; APPLICATION NUMBER: US/08/779,801
24 ; FILING DATE: January 7, 1997

```


[illegible]

```

Db 3162 AGGATGTGAGAGCGGGTCCAGGCCGCCGCCGAGGAGTCCCCCAGGGGCTGTCCACAC 3221
QY 585 AGAAANTCTCGGGGCTGTGGGGTGCNCCCTGGCAACAGCCGGGGTCCCAAGCCCCACCG 643
Db 3222 GCGGGGGCCGGTGGCCGCATCTGGCCCGCAGCTGTCCAGGGGCCCGGGTCCCG 3280

RESULT 12
US-08-531-927B-3
; Sequence 3, Application US/08531927B
; Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491
; TITLE OF INVENTION: Disease Gene and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/531,927B
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP H6-251600
; FILING DATE: 21-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: ATE95-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 145
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 194
US-08-531-927B-3

Query Match 7.0%; Score 45.8; DB 2; Length 325;
Best Local Similarity 59.4%; 2; Mismatches 24; Indels 0; Gaps 0;
Matches 59; Conservative

QY 332 TGAAGCATCTCAAAAGGAGGAATAGTCAAAGCAGCAGCGGGCGCGCGCGGCGACGA 391
Db 115 TGAATGTTTCAGACAGCAGCAAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 174
QY 392 GCACGACGACGAGGAGTGGGGGCC 416
Db 175 GCACGACGACGACGACGAGSGGGAC 199

RESULT 13
US-09-249-585A-2

```

```
; REFERENCE DOCKET NUMBER A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

Query Match          6.%; Score 45.4; DB 4; Length 2580;
Best Local Similarity 49.0%; Pred. No. 0.0071;
Matches 142; Conservative 1; Mismatches 145; Indels 2; Gaps 1

QY   346 AGGAGGAATAGTCAAAAGCAGCAGCGGGCGGGCGGGCAGCAGCAGCAGCAGCAGG 405
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    896 AGCAGAGGGCGAGGCAGGACGAGGAGGGCAGAGGGCCAGGAGCAGAGNGNAGGGCCAGG 955

QY   406 AGGTGGGGGCCCTCTGCAGGTACTCCGGGCGGGCGGACGACCGAGAGTGTCACAGTCCCOCGG 465
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    956 AGGGCGCAGGAGGGCGAGGACGACGAGGAGGAGGGCGGAGGACGAGGAGGGCGCAGGAGGGCAGG 1015

QY   466 GAGGCCACCTTTCCCTGGAGTGCTGTAGACAGAGGGAAGGAGGAGGCCAGCAGCAGGAA 525
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1016 AGGGCGAGA--GCAGAGGGCGCAGCAGCAGGAGAGGGGCAGAGGGCGCAGGAGGGCGCA 1073

QY   526 TCAGAGCGAGGCAAAGGGCGGCGAGGAACGTANGAAGATGACSCGGGAGGCGCGCGGAAA 585
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1074 GGAGCAGGAGGGCGCAGGACGAGGAGGGCGCAGGACGAGGAGGGCGCAGGACGAGGAGGCGCA 1133

QY   586 GAAGATCTCGGGGCTGTGGGGGTGTCNCCTTGGCACGACGCCGGGTGTCCTCCAAG 635
      | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    1134 GGAGGGCGCAGGAGGAGGGCGGAGGGGGCGAGGAGGGGCGAGGACGAGGGGGCGCAGG 1183
```

Search completed: January 10, 2003, 06:37:32
Job time : 72.2338 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 01:01:36 ; Search time 85.8611 Seconds
(without alignments)
3362.497 Million cell updates/sec

Title: US-09-581-500b-12

Perfect score: 656

Sequence: 1 gccacaacaaataaataat.....cccaccgcgagacccgcgga 656

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications_NA.*
- 1: /cgn2.6/prodata/2/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2.6/prodata/2/pubpna/ECT_NEW_PUB.seq.*
 - 3: /cgn2.6/prodata/2/pubpna/US06_NEW_PUB.seq.*
 - 4: /cgn2.6/prodata/2/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2.6/prodata/2/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2.6/prodata/2/pubpna/ECTUS_PUBCOMB.seq.*
 - 7: /cgn2.6/prodata/2/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2.6/prodata/2/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2.6/prodata/2/pubpna/US09_NEW_PUB.seq.*
 - 10: /cgn2.6/prodata/2/pubpna/US09_PUBCOMB.seq.*
 - 11: /cgn2.6/prodata/2/pubpna/US10_NEW_PUB.seq.*
 - 12: /cgn2.6/prodata/2/pubpna/US10_PUBCOMB.seq.*
 - 13: /cgn2.6/prodata/2/pubpna/US60_NEW_PUB.seq.*
 - 14: /cgn2.6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	91.8	14.0	1476	9	US-09-974-298-94
2	47.8	7.3	2724	12	US-10-044-090-309
3	46.8	7.1	5222	10	US-09-416-384A-1
4	46.8	7.1	5566	10	US-09-416-384A-4
5	45.6	7.0	599	10	US-09-864-761-7244
6	45.6	7.0	599	10	US-09-864-761-5950
7	45.4	6.9	3886	9	US-10-124-800-37
8	44.3	6.8	230	10	US-09-864-761-23975
9	44	6.7	210	10	US-09-864-761-22135
10	44	6.7	293	10	US-09-864-761-18923
11	44	6.7	459	10	US-09-864-761-2182
12	44	6.7	464	10	US-09-864-761-5361
13	43	6.6	527	10	US-09-925-297-272
14	42.8	6.5	311	10	US-09-960-352-14967
15	42.6	6.5	2453	12	US-10-005-858-1
16	42.4	6.5	612	10	US-09-878-574-4578
17	42.2	6.4	1310	9	US-09-849-243-13
18	42.2	6.4	3263	9	US-09-849-243-15
19	42.2	6.4	4286	9	US-09-849-243-14

c	20	42	6.4	1511	9	US-10-078-650-11	Sequence 11, Appl
c	21	42	6.4	3641	9	US-10-078-650-1	Sequence 1, Appl
	22	41.6	6.3	226	10	US-09-728-445-159	Sequence 159, App
	23	41.6	6.3	2572	10	US-09-925-300-486	Sequence 486, App
	24	41.6	6.3	3715	10	US-09-880-107-2300	Sequence 2300, Ap
c	25	41.4	6.3	501	10	US-09-864-761-8824	Sequence 8824, Ap
c	26	41.4	6.3	2336	12	US-10-044-090-212	Sequence 212, App
c	27	41.4	6.3	4281	12	US-10-002-600-23	Sequence 23, Appl
c	28	41.2	6.3	122	10	US-09-864-761-25491	Sequence 25491, A
c	29	41.2	6.3	1518	10	US-09-976-165-20	Sequence 20, Appl
c	30	41.2	6.3	2636	10	US-09-976-165-21	Sequence 21, Appl
c	31	41	6.2	441	10	US-09-893-737-41	Sequence 41, Appl
c	32	41	6.2	537	10	US-09-919-580-475	Sequence 475, App
c	33	41	6.2	572	10	US-09-919-580-570	Sequence 570, App
c	34	41	6.2	1064	10	US-09-804-682-29	Sequence 29, Appl
c	35	40.8	6.2	1160	9	US-09-946-807-108	Sequence 108, App
c	36	40.8	6.2	1160	10	US-09-795-668-108	Sequence 108, App
c	37	40.8	6.2	1160	10	US-09-795-668-108	Sequence 108, App
c	38	40.8	6.2	1503841	9	US-09-946-807-1	Sequence 1, Appl
c	39	40.8	6.2	1503841	10	US-09-795-668-1	Sequence 1, Appl
c	40	40.6	6.2	401	9	US-09-946-807-1097	Sequence 1097, Ap
c	41	40.6	6.2	401	10	US-09-795-668-1097	Sequence 1097, Ap
c	42	40.6	6.2	401	10	US-09-795-668-1097	Sequence 1097, Ap
c	43	40.6	6.2	1362	10	US-09-822-830A-19	Sequence 19, Appl
c	44	40.6	6.2	1362	10	US-09-822-830A-19	Sequence 19, Appl
c	45	40.6	6.2	1636	10	US-09-954-456-1185	Sequence 1185, Ap

ALIGNMENTS

RESULT 1
US-09-974-298-94
; Sequence 94, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 94
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 589880CBI
US-09-974-298-94

Query Match 14.08; Score 91.8; DB 9; Length 1476;
Best Local Similarity 93.28; Pred. No. 4.6e-15;
Matches 96; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 315 CTCATCAAGCGCGAGCTTACAAAGGAGGAGTACAAAGCAGCAGCGCGG 374
Db 1 CTCATCAAGCGCGAGCTTACAAAGGAGGAGTACAAAGCAGCAGCGCGG 60
QY 375 CGCGCGCGCGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 417
Db 61 CGCGCGCGCGCGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 103
RESULT 2
US-10-044-090-309/c
; Sequence 309, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman

```

; LOCATION: 4872
; OTHER INFORMATION: 8-58-301 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 3606
; OTHER INFORMATION: insertion of AGAG in SEQID4
; NAME/KEY: primer_bind
; LOCATION: 4572..4587
; OTHER INFORMATION: 8-58.pu
; NAME/KEY: primer_bind
; LOCATION: 4990..5005
; OTHER INFORMATION: 8-58.rp complement
; NAME/KEY: misc_binding
; LOCATION: 4849..4895
; OTHER INFORMATION: 8-58-301.probe
; NAME/KEY: primer_bind
; LOCATION: 4853..4871
; OTHER INFORMATION: 8-58-301.mis
; NAME/KEY: primer_bind
; LOCATION: 4873..4891
; OTHER INFORMATION: 8-58-301.mis complement
; NAME/KEY: misc_feature
; LOCATION: 148,686,902,1258,1322,2440,2794,2852,3018..3023,3052..3053,3064
; OTHER INFORMATION: n=a, g, c or t
US-09-416-384A-1

Query Match          7.1%; Score 46.8; DB 10; Length 5222;
Best Local Similarity 58.7%; Pred. No. 0.0074;
Matches 81; Conservative 0; Mismatches 57; Indels 0; Gaps 0

QY   346 AGGAGGAATGTCACAAAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGAGCACAGCAGCAGCAGG 405
Db    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dd    3964 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 403
QY   406 AGTGGGGGCCTCTCCAGGTACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 465
Db    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dd    4024 AGTAGGCCCTCTCTGGCCCCCGCTCTTGCGAGCATGGGGGAGTCTCTGCCGCCGCCAGG 4083
QY   466 GAGGCCACCTCTCTCCCTG 483
Db    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dd    4084 CACAGACTCTCTCTCG 4101

RESULT 4
US-09-416-384A-4
; Sequence 4, Application US/09416384A
; Patent No. US20020081584A1
; GENERAL INFORMATION:
; APPLICANT: BLUMENFELD, Marta
; APPLICANT: BOUGUELERET, Lydie
; APPLICANT: CHUMAKOV, Ilya
; APPLICANT: COHEN, Daniel
; APPLICANT: ESSIUX, Laurent
; TITLE OF INVENTION: Genes, proteins and biallelic markers related to cent
; FILE REFERENCE: GENSET.045AUS
; CURRENT FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: US/09/416,384A
; PRIOR APPLICATION NUMBER: 60/106,457
; PRIOR FILING DATE: 1999-10-30
; PRIOR APPLICATION NUMBER: 60/103,955
; PRIOR FILING DATE: 1998-10-12
; PRIOR APPLICATION NUMBER: 60/132,277
; PRIOR FILING DATE: 1999-05-03
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patent.pm
; SEQ ID NO 4
; LENGTH: 5566
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 659..2032
; NAME/KEY: allele
; LOCATION: 4484

```


Qy	361	AGCAGACGGCGCGCGCGCGCGCGAGCAGCAGCAGCAGGAGTGTGGGGCCCTCTG	420
Db	313	AGCAGAGTATCGGCGCGCGCGCGCGCAGCAGCAGCAGCAGCAGGAGAAAG	372
Qy	421	CCAGGTACCGCGCGCGCGCAGCAGGAGTGCCTAGTTCGCCGGGAGGCACCTCTTTC	480
Db	373	ACCAGGACGAGTTGCACGTGCTCCCTCCCTCCCTCCGCTCAGGCCAGGAGTGC	432
Qy	481	CTGGAGTGCTGAGACAGGGGAGGGAGGAGCCAGCAGCAGGAATCATCAGCGAGGCAAA	540
Db	433	TGGGACCTGTGGTCTGGCAGTGGGAGCAGGGGAGCGCTGTCACGACAGGGTGGGAA	492

[illegible]

RESULT 9
US-09-864-761-22135

AC0091.13
 US-09-925-297-272
 ; Sequence 272, Application US-09925297
 ; Patent No. US20020081659A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, P
 ; FILE REFERENCE: PA105
 ; CURRENT APPLICATION NUMBER: US/09/925
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/0
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 50/124, 270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 928
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO: 272

```

Db      139  TCAAGGGCGAGAACGGCGCGCGGGAACAAACAGAGCTGGAGAGGAACAGNNGCAG 198
      542  GCGGCGAGGAA 552
      199  AAGGNGANGGA 209
      199  AAGGNGANGGA 209

RESULT 15
US-10-005-858-1
; Sequence 1, Application US/10005858
; Patent No. US20020116729A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING NTP1
; TITLE OF INVENTION: PHOSPHATASE GENE DISRUPTIONS
; FILE REFERENCE: R-690
; CURRENT APPLICATION NUMBER: US/10/005,858
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 60/251,802
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2453
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-005-858-1

Query Match          6.5%; Score 42.6; DB 12; Length 2453;
Best Local Similarity 59.5%; Pred. No. 0.07; Indels 0; Gaps
Matches 72; Conservative 0; Mismatches 49;

QY      346  AGAGGAGAAATAGTCTAAAGCAGCAGCGCGCGCGCGCGCGCGAGCAGCAGCAGCAGCAGG 405
Db      1826  AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1885
QY      406  AGGTGGGGGCGCTCTGCCAGTACCGGGGGGGGGCAGGCGAGGAGTGCCAGATTCCCGG 465
Db      1886  AGTAGTAGTAGTAGTGTACCTGCCGAGGCGGGATGTGGCGACCGCGGTGCCCGGAGGAGCTGCT 1945
QY      466  G 466
Db      1946  G 1946

Search completed: January 10, 2003, 06:43:11
Job time : 98.8611 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 21:14:32 ; Search time 2221.92 Seconds
(without alignments)
4781.561 Million cell updates/sec

Title: US-09-581-500b-12
Perfect score: 656
Sequence: 1 gccacacacacaaatgaaat.....ccacacgcgagaccocgcga 656

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estma.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_esti.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pin.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	384.4	58.6	636	14 BQ807824	BQ807824 NISC_Kk10
2	142.8	21.8	930	14 BQ435307	BQ435307 AGENCOURT
3	137.4	20.9	270	9 AA904435	AA904435 ok07q11.s
4	126.4	19.3	244	9 AA889105	AA889105 am3sh04.s
5	107	16.3	601	9 AL708543	AL708543 DKFZp686B
6	91.8	14.0	707	13 B1819086	B1819086 603033444

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

7	76.4	11.6	742	13	B1756040
8	70	10.7	896	12	BG499418
9	67	10.2	925	14	BQ929418
c 10	63.2	9.6	925	17	CNS0091P
11	61.4	9.4	839	17	CNS004NB
c 12	61.4	9.4	1071	17	CNS000EM
c 13	56.2	8.6	844	17	CNS0052P
c 14	54.8	8.4	348	9	AL664001
15	54.6	8.3	691	13	B1547708
16	54.4	8.3	300	9	AA853497
c 17	54.4	8.3	456	9	AL554951
18	54	8.2	876	17	CNS03C63
19	53.8	8.2	1052	13	BM553599
c 20	53.8	8.2	1336	14	BM810024
c 21	53.6	8.2	1101	17	CNS0153F
c 22	53.2	8.1	764	17	AG060189
c 23	53.2	8.1	1101	17	CNS00KXY
24	53.2	8.1	1145	10	B8422200
25	53	8.1	548	9	AL597104
26	53	8.1	897	17	AG060530
c 27	53	8.1	912	17	CNS006N3
c 28	52.8	8.0	609	10	AV704516
c 29	52.8	8.0	773	13	B1892255
c 30	52.8	8.0	977	9	AL578905
c 31	52.8	8.0	1199	14	BQ706827
32	52.6	8.0	613	17	AG043036
c 33	52.6	8.0	750	17	CNS04PWH
34	52.2	8.0	677	17	AG154064
35	52.2	8.0	949	17	AG171092
36	52	7.9	925	12	BG742444
c 37	51.8	7.9	430	12	BG55144
c 38	51.6	7.9	846	17	CNS00KOV
39	51.6	7.9	949	14	BQ940345
c 40	51.6	7.9	967	17	CNS01LMA
41	51.6	7.9	1036	17	CNS010BS
42	51.4	7.8	533	9	AL572608
c 43	51.4	7.8	798	13	B1951298
44	51.4	7.8	834	17	AG131468
45	51.2	7.8	450	9	AL513867

ALIGNMENTS

RESULT 1
BQ807824/c 636 bp mRNA linear EST 31-JUL-2002
LOCUS NISC_Kk10f04.y1 NCI_CGAP_Brn72 Macaca mulatta cDNA clone
DEFINITION IMAGE:5331199.5', mRNA sequence.
ACCESSION BQ807824
VERSION BQ807824.1 GI:22032033
KEYWORDS EST.
SOURCE rhesus monkey.
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciae; Macaca.
REFERENCE 1 (bases 1 to 636)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
CDNA Library Preparation: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Plate: LLAM11840 row: K column: 8
Seq primer: M13Rpl reverse primer (ABI).

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L1AM13511 row: p column: 22
 High quality sequence stop: 607.
 Location/Qualifiers
 1..930
 /organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:6160773"
/clone_lib="NIH_MGC_72"

```

```
/tissue_type="melanotic melanoma"  
/lab_host="DH10B (phage-resistant)"
```

Site_1: NotI
Site_2: SalI; Cloned unidirectionally. Primer: Oligo
Average insert size 2 kb. Library constructed by Vi-

Average insert size kb.	Library constructed by
268 a	210 c
263 a	186 t
3 others	

Length 21.8%; Score 142.8; DB 14; Length 930;
Similarity 98.6%; Pred. No. 6.1e-25;

1444; conservative 0; mismatches 2; indels 0; gaps

CCGCAAGCCAGAGCAATCGAACGGTCTGAGTCATCTGCCGGAAGCCTTGCCCTCAATC 79

AGGGGGACGTGAAGCATCTACAAAGGAGGAATAGTCAAAGCAGCAGCGGGGGGGC 381

AGCGGACGTGAAGCATCTACAAGGAGGAATAGTCAAGCAGCGGGCGGGCG 139

SCGGCAGCAGCAGCAGCAGCAGGAG 407

GCGGCAGCAGCAGCAGCAGGAG 165

AA904435 270 bp mRNA linear EST 09-JUN-2007

OX07911: si Soares_NF1-1_GBC_31 Homo sapiens CDNA Clone IMAGE:1507172 3', mRNA sequence.
AA904435

AA904435.1 GI:3039558
EST.

human.
Homo sapiens

1 (bases 1 to 270)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Vertebrata; Vertebrata; Homo.

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP)

**Tumor Gene Index
Unpublished (1997)**

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through EMLL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information
 Insert Length: 437 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 232.

Location/Qualifiers
1. .270

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/accession="F001170"
```

```
/clone="IMAGE:150/1/2"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="ny10B"
```

DATA - 3 SOURCE /

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19w, testis NHT, and B-cell NCI CGAP GC81) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

```

BASE COUNT      77 a 66 c 74 g 53 t
ORIGIN
Query Match      20.9%; Score 137.4; DB 9; Length 270;
Best Local Similarity 99.3%; Pred. No. 1.2e-23;
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 162 CAATGTGTACACTTTTATATAAAACACCCCTCCAAAGGACGAGCACTGGCCCTCTCTCCG 221
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 5 CAATGTGTACACTTTTATATAAAACACCCCTCCAAAGGACGAGCACTGGCCCTCTCTCCG 64

QY 222 GTGCCACAGACATCCACACAGCCCAAGAATCAGGGATTGCACAAAGCCAGAGCAATCG 281
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 65 GTGCCCGCAGACATCCACACAGCGCCCAAGAATCAGGGATTGCACAAAGCCAGAGCAATCG 124

QY 282 AACGGTCTGAGTCATCTG 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 125 AACGGTCTGAGTCATCTG 143

RESULT 4
AA889105
LOCUS      AA889105      244 bp      mRNA      linear      EST 04-JAN-1999
DEFINITION IMAGE:1471063 3', mRNA sequence.
ACCESSION  AA889105
VERSION    AA889105.1 GI:3015984
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 244)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL   Tumor Gene Index
COMMENT    Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cga@r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 482 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 192.
FEATURES
Location/Qualifiers
1..244
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1471063"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19w, testis NHT, and B-cell
NCI CGAP GC81) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,

```

```

726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.
BASE COUNT      74 a 61 c 63 g 46 t
ORIGIN
Query Match      19.3%; Score 126.4; DB 9; Length 244;
Best Local Similarity 98.6%; Pred. No. 6.4e-21;
Matches 138; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 162 CAATGTGTACACTTTTATATAAAACACCCCTCCAAAGGACGAGCACTGGCCCTCTCTCCG 221
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 4 CAATGTGTACACTTTTATATAAAACACCCCTCCAAAGGACGAGCACTGGCCCTCTCTCCG 63

QY 222 GTGCCACAGACATCCACACAG-GCCCAAGAATCAGGGATTGCACAAAGCCAGAGCAATC 280
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 64 GTGCCCGCAGACATCCACACAGGCCCAAGAATCAGGGATTGCACAAAGCCAGAGCAATC 123

QY 281 GAACGGTCTGAGTCATCTG 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 124 GAACGGTCTGAGTCATCTG 143

RESULT 5
AL708543      601 bp      mRNA      linear      EST 22-MAR-2002
LOCUS      DKFP686B1453_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION DKFP686B1453 5', mRNA sequence.
ACCESSION  AL708543
VERSION    AL708543.1 GI:19691898
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 601)
AUTHORS   Wambutt, R., Heubner, D., Mewes, W., Weil, B. and Wiemann, S.
TITLE     EST (Wambutt, R., Heubner, D., Mewes, H.W., Weil, B. and Wiemann, S.)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Wambutt R
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKFP686B1453) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Location/Qualifiers
1..601
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFP686B1453"
/clone_lib="686 (synonym: hlcc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplEx2; Site_1: SfiI; Site_2: SfiI;
cDNA-collection"
BASE COUNT      167 a 133 c 167 g 134 t
ORIGIN
Query Match      16.3%; Score 107; DB 9; Length 601;
Best Local Similarity 100.0%; Pred. No. 4.4e-16;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 318 AATCAAGCGGACGTGAAGCATCTCAAAAGGAGGATAGTCAAAAGCAGAGCGGGCGG 377
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 AATCAAGCGGACGTGAAGCATCTCAAAAGGAGGATAGTCAAAAGCAGAGCGGGCGG 60

QY 378 CGGCGGCGCAGCAGCAGCAGCAGGAGGTGGGGGCGCTCTGCCAG 424

```


|||||

Db 61 CGCGCGCGCAGCAGCAGCAGGAGTGGGGCCCTCTGCCAG 107

RESULT 6

BI1819086 707 bp mRNA linear EST 04-OCT-2001

LOCUS 603033444F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5174458 5',

DEFINITION mRNA sequence.

ACCESSION BI1819086

VERSION BI1819086.1 GI:15930636

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 707)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLML1434 row: h column: 11

High quality sequence stop: 697.

Location/Qualifiers

1..707

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5174458"

/clone_lib="NIH_MGC_115"

/lab_host="DH10B"

/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 Kb, insert size range 1-3 Kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."

BASE COUNT 208 a 149 c 191 g 159 t

ORIGIN

Query Match 14.0%; Score 91.8; DB 13; Length 707;

Best Local Similarity 97.9%; Pred. No. 2.6e-12;

Matches 93; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 330 CCGTGAAGCATCTCAAGAGGAGGATAGTCAAGCAGCAGCGCGCGCGCGCGCGCAG 389

Db 1 CCGTGAAGCATCTCAAGAGGAGGATAGTCAAGCAGCAGCGCGCGCGCGCGCGCAG 60

QY 390 CAGCAGCAGCAGCAGGAGTGGGGCCCTCTGCCAG 424

Db 61 CAGCAGCAGCAGCAGGAGTGGGGCCCTCTGCCAG 95

RESULT 7

BI1756040 742 bp mRNA linear EST 25-SEP-2001

LOCUS 603030386F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200677 5',

DEFINITION mRNA sequence.

ACCESSION BI1756040

VERSION BI1756040.1 GI:15747618

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 896)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTF

CDNA Library Preparation: CLONETECH Laboratories, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLML1502 row: 1 column: 22

High quality sequence stop: 742.

Location/Qualifiers

1..742

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5200677"

/clone_lib="NIH_MGC_114"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 Kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH_MGC Library."

BASE COUNT 223 a 153 c 200 g 166 t

ORIGIN

Query Match 11.6%; Score 76.4; DB 13; Length 742;

Best Local Similarity 98.7%; Pred. No. 1.7e-08;

Matches 77; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 330 CCGTGAAGCATCTCAAGAGGAGGATAGTCAAGCAGCAGCGCGCGCGCGCGCGCAG 389

Db 1 CCGTGAAGCATCTCAAGAGGAGGATAGTCAAGCAGCAGCGCGCGCGCGCGCGCAG 60

QY 390 CAGCAGCAGCAGCAGGAG 407

Db 61 CAGCAGCAGCAGCAGGAG 78

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 742)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLML1502 row: 1 column: 22

High quality sequence stop: 742.

Location/Qualifiers

1..742

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5200677"

/clone_lib="NIH_MGC_114"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 Kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH_MGC Library."

BASE COUNT 223 a 153 c 200 g 166 t

ORIGIN

Query Match 11.6%; Score 76.4; DB 13; Length 742;

Best Local Similarity 98.7%; Pred. No. 1.7e-08;

Matches 77; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 330 CCGTGAAGCATCTCAAGAGGAGGATAGTCAAGCAGCAGCGCGCGCGCGCGCGCAG 389

Db 1 CCGTGAAGCATCTCAAGAGGAGGATAGTCAAGCAGCAGCGCGCGCGCGCGCGCAG 60

QY 390 CAGCAGCAGCAGCAGGAG 407

Db 61 CAGCAGCAGCAGCAGGAG 78

RESULT 8

BI1756040 896 bp mRNA linear EST 27-MAR-2001

LOCUS 602546648F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4669282 5',

DEFINITION mRNA sequence.

ACCESSION BI1756040

VERSION BI1756040.1 GI:13460947

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 896)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTF

CDNA Library Preparation: CLONETECH Laboratories, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

[illegible]

```

QY      535  GGCAGAGGGGGCAGGA  551
          | | | | | | | | | |
Db      319  TGGCAGGGCAGGCAGA  335

Search completed: January 10, 2003, 06:15:15
Job time : 2229.02 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 9, 2003, 19:57:20 ; Search time 83.435 Seconds
(without alignments)
7673.783 Million cell updates/sec

Title: US-09-581-500B-13
Perfect score: 22
Sequence: 1 atcgaacgggttcgtgcatot 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank:
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	22	100.0	22	6	AX021007	AX021007 Sequence
2	22	100.0	656	6	AX021006	AX021006 Sequence
3	22	100.0	191395	9	AC009802	AC009802 Homo sapi
4	22	100.0	191793	2	AP001897	AP001897 Homo sapi
5	22	100.0	198291	2	AC067875	AC067875 Homo sapi
6	22	100.0	201734	2	AC090224	AC090224 Homo sapi
7	20.4	92.7	186351	2	AP001503	AP001503 Homo sapi
8	17.8	80.9	95448	9	AC005572	AC005572 Homo sapi
9	17.8	80.9	124092	2	CNS08C9K	AL732645 Oryza sat
10	17.8	80.9	175238	10	AL060903	AL606903 Mouse DNA
11	17.8	80.9	176641	2	AC084869	AC084869 Homo sapi
12	17.8	80.9	178103	2	AC118856	AC118856 Rattus no
13	17.8	80.9	228664	2	AL606971	AL606971 Mus muscu
14	17.8	80.9	258929	9	AC010433	AC010433 Homo sapi
15	17.4	79.1	192082	10	AL591490	AL591490 Mouse DNA
16	17.2	78.2	1163	8	AY088527	AY088527 Arabidops
17	17.2	78.2	1175	8	ATRNASC5D	X90454 A.thaliana
18	17.2	78.2	1689	6	A75661	A75661 Sequence 6
19	17.2	78.2	2294	8	AF105034	AF105034 Arabidops
20	17.2	78.2	3004	8	AF069468	AF069468 Arabidops
21	17.2	78.2	3084	6	A75674	A75674 Sequence 19
22	17.2	78.2	3569	3	HC024941	AJ245941 haemochu
23	17.2	78.2	10174	9	HS598F21A	AL096857 Novel hum
24	17.2	78.2	13202	6	AX346386	AX346386 Sequence
25	17.2	78.2	13202	6	AX346387	AX346387 Sequence
26	17.2	78.2	99886	9	HS598F2	AL021579 Human DNA
27	17.2	78.2	103904	8	ATAC021640	AC021640 Arabidops
28	17.2	78.2	119670	2	AP003988	AP003988 Oryza sat
29	17.2	78.2	121622	2	AC124971	AC124971 Medicago
30	17.2	78.2	134282	2	AP004229	AP004229 Oryza sat
31	17.2	78.2	156302	2	AC123298	AC123298 Rattus no
32	16.8	76.4	1356	1	AF005745	AF005745 Unculture
33	16.8	76.4	55784	2	AC120157	AC120157 Mus muscu
34	16.8	76.4	58376	2	AC127705	AC127705 Rattus no
35	16.8	76.4	71290	2	AC123313	AC123313 Rattus no
36	16.8	76.4	187392	10	AL683890	AL683890 Mouse DNA
37	16.8	76.4	207358	2	AL732494	AL732494 Mus muscu
38	16.8	76.4	214225	2	AC094155	AC094155 Rattus no
39	16.4	74.5	41084	3	AC006603	AC006603 Caenorhab
40	16.4	74.5	46855	8	AC078898	AC078898 Arabidops
41	16.4	74.5	98890	2	AC130896	AC130896 Homo sapi
42	16.4	74.5	166458	2	AC084737	AC084737 Homo sapi
43	16.4	74.5	178012	2	AC131214	AC131214 Homo sapi
44	16.4	74.5	191150	2	AC129402	AC129402 Rattus no
45	16.4	74.5	263876	9	AC019233	AC019233 Homo sapi

ALIGNMENTS

RESULT 1
AX021007
LOCUS
DEFINITION Sequence 13 from Patent WO9932643.
ACCESSION AX021007
VERSION AX021007.1 GI:10044670
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 22)
AUTHORS Del-Favero,J., Raeymaekers,P. and Van Broeckhoven,C.
TITLE Mood disorder gene
JOURNAL Patent: WO 9932643-A 13 01-JUL-1999;

AX021007 22 bp DNA linear PAT 07-SEP-2000

REFERENCE

2. (bases 1 to 191395)

AUTHORS

Birren B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barua, N., Beckerly, R., Benn, J., Brown, A., Castie, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., deArillano, K., Depayre, E., Devlin, K., Dewar, K., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardvany, S., Gilbert, D., Grant, G.,

```

repeat_region /note="Single-stranded terminator coverage."
repeat_region 2208..2215 /rpt_family="(TATATG)n"
repeat_region 2594..2648 /rpt_family="LIME"
repeat_region complement(2649..2991) /rpt_family="MER2"
repeat_region 2992..3475 /rpt_family="LIME"
repeat_region complement(4330..4437) /rpt_family="FRAM"
repeat_region 4438..4567 /rpt_family="(TA)n"
repeat_region 4599..4959 /rpt_family="AT_rich"
repeat_region 4982..5151 /rpt_family="(TATG)n"
repeat_region 5144..5311 /rpt_family="(TATA)n"
repeat_region 5257..5441 /rpt_family="(TA)n"
repeat_region 5444..5525 /rpt_family="AT_rich"
repeat_region 5533..5708 /rpt_family="(TTATA)n"
repeat_region 5742..5907 /rpt_family="(TTATA)n"
repeat_region complement(5966..6047) /rpt_family="FRAM/PRAM"
repeat_region 6682..6703 /rpt_family="AT_rich"
repeat_region 7034..7933 /rpt_family="L1P16"
repeat_region 7935..8006 /rpt_family="AluY"
repeat_region 8009..9111 /rpt_family="L1P16"
repeat_region 9112..9165 /rpt_family="AT_rich"
repeat_region 9216..9496 /rpt_family="AluX"
repeat_region 9497..9520 /rpt_family="(CAAA)n"
repeat_region 9961..10264 /rpt_family="AluY"
repeat_region complement(10837..10840) /note="Single-stranded terminator coverage."
repeat_region complement(11156..11515) /rpt_family="THEIC"
repeat_region 11856..11987 /rpt_family="MIR"
repeat_region 12592..12643 /rpt_family="L2"
repeat_region 13683..13994 /rpt_family="AluX"
repeat_region complement(13996..14160) /rpt_family="L1MEC"
repeat_region 14788..15255 /rpt_family="MER67D"
repeat_region complement(15875..15996) /rpt_family="MER5B"
repeat_region 16101..16135 /rpt_family="AT_rich"
repeat_region 16136..16304 /rpt_family="L1M4"
repeat_region complement(16572..16647) /rpt_family="MAE51"
repeat_region complement(16716..17011) /rpt_family="AluY"
repeat_region 17045..17780 /rpt_family="L1ME3"
repeat_region 17827..17938 /rpt_family="L1ME3"

```

```

repeat_region complement(17945..18133) /rpt_family="MER58A"
repeat_region 18222..18755 /rpt_family="LIME3"
repeat_region 18843..19019 /rpt_family="LIME3"
repeat_region 19573..19616 /rpt_family="(CACG)n"
repeat_region 20506..20544 /rpt_family="AT_rich"
repeat_region 21894..22184 /rpt_family="AluSq"
repeat_region complement(22787..22998) /rpt_family="LFR40a"
repeat_region complement(23074..23133) /rpt_family="LFR40b"
repeat_region 23184..23540 /rpt_family="L2"
repeat_region complement(23675..25225) /rpt_family="pTR5"
repeat_region 25229..25582 /rpt_family="THELB"
repeat_region 25879..26031 /rpt_family="MLT1A1"
repeat_region 26032..26321 /rpt_family="AluX"
repeat_region 26322..26531 /rpt_family="MLT1A1"
repeat_region 26623..26736 /rpt_family="L2"
repeat_region 26943..27439 /rpt_family="L2"
repeat_region 27914..27985 /rpt_family="MIR"

```

Query Match 100.0%; Score 22; DB 9; Length 191395;
 Best Local Similarity 100.0%; Pred. No. 0.37;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCGACGGTTCGATCATCT 22
 |||||

Db 97997 ATCGACGGTTCGATCATCT 97976

RESULT 4

AP001897 191793 bp DNA linear HTG 30-MAY-2000
 LOCUS Homo sapiens chromosome 18 clone Rp11-693A18 map 18q22, WORKING
 DEFINITION DRAFT SEQUENCE, 54 unordered pieces.
 ACCESSION AP001897
 VERSION AP001897.2 GI:8117548
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens DNA, clone:Rp11-693A18.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 191793)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens 191,793 genomic DNA of 18q22

JOURNAL Published Only in Database (2000)

REFERENCE 2 (bases 1 to 191793)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (24-APR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

Japan (E-mail:hattori@gsr.riken.go.jp,

URL:http://hqp.gsc.riken.go.jp/, Tel:81-42-778-9923,

Fax:81-42-778-9924)

COMMENT On May 30, 2000 this sequence version replaced gi:7649784.

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: <http://ngp.gsc.riken.go.jp/>

Contact: hattori@gsc.riken.go.jp

----- Project Information

Center project name: HumDraft18

Center clone name: Rp11-693A18

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 167461 bases at least Q40

Consensus quality: 174712 bases at least Q30

Consensus quality: 180769 bases at least Q20

Insert size: 186493; sum-of-contigs

Quality coverage: 4.60x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 54 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 10723 contig of 10723 bp in length
10824 19450 contig of 8627 bp in length
19551 28739 contig of 9189 bp in length
28840 37796 contig of 8957 bp in length
37897 45643 contig of 7747 bp in length
45744 53200 contig of 7457 bp in length
53301 59466 contig of 6166 bp in length
59467 65023 contig of 5457 bp in length
65024 71725 contig of 6602 bp in length
71726 77046 contig of 5221 bp in length
77047 82059 contig of 4913 bp in length
82160 86516 contig of 4356 bp in length
86616 90337 contig of 4221 bp in length
90338 99276 contig of 3919 bp in length
99377 103751 contig of 4375 bp in length
103852 108407 contig of 4556 bp in length
108508 112468 contig of 3961 bp in length
112569 116411 contig of 3843 bp in length
116512 120690 contig of 4179 bp in length
120791 125789 contig of 4999 bp in length
125890 129429 contig of 3540 bp in length
129530 132875 contig of 3346 bp in length
132976 136718 contig of 3743 bp in length
136819 140204 contig of 3386 bp in length
140305 143313 contig of 2908 bp in length
143313 146311 contig of 2999 bp in length
146412 148694 contig of 2283 bp in length
148795 151551 contig of 2757 bp in length
151652 153471 contig of 1820 bp in length
153572 155905 contig of 2334 bp in length
156006 157930 contig of 1925 bp in length
158031 160335 contig of 2305 bp in length
160436 161957 contig of 1522 bp in length
162058 164115 contig of 2058 bp in length
164215 166374 contig of 2159 bp in length
164735 168634 contig of 2160 bp in length
168735 170796 contig of 2062 bp in length
170897 172731 contig of 1835 bp in length
172832 174464 contig of 1633 bp in length
174565 175680 contig of 1116 bp in length
175781 176918 contig of 1138 bp in length
177019 178882 contig of 1804 bp in length
178923 180077 contig of 1155 bp in length
180178 181248 contig of 1071 bp in length
181349 182499 contig of 1151 bp in length
182600 182806 contig of 207 bp in length
182907 183992 contig of 1086 bp in length
184093 185198 contig of 1106 bp in length

185299 186320 contig of 1022 bp in length
186421 187601 contig of 1181 bp in length
187702 189099 contig of 1398 bp in length
189200 190576 contig of 1377 bp in length
190677 191793 contig of 1117 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 10723: contig of 10723 bp in length
10724 10823: gap of 100 bp
10824 19450: contig of 8627 bp in length
19451 19550: gap of 100 bp
19551 28739: contig of 9189 bp in length
28740 28839: gap of 100 bp
28840 37796: contig of 8957 bp in length
37797 37896: gap of 100 bp
37897 45643: contig of 7747 bp in length
45644 45743: gap of 100 bp
45744 53200: contig of 7457 bp in length
53201 53300: gap of 100 bp
53301 59466: contig of 6166 bp in length
59467 59566: gap of 100 bp
59567 65023: contig of 5457 bp in length
65024 65123: gap of 100 bp
65124 71725: contig of 6602 bp in length
71726 71825: gap of 100 bp
71826 77046: contig of 5221 bp in length
77047 77146: gap of 100 bp
77147 82059: contig of 4913 bp in length
82060 82159: gap of 100 bp
82160 86515: contig of 4356 bp in length
86516 86615: gap of 100 bp
86616 90836: contig of 4221 bp in length
90837 90936: gap of 100 bp
90937 95257: contig of 4321 bp in length
95258 95357: gap of 100 bp
95358 99276: contig of 3919 bp in length
99277 99376: gap of 100 bp
99377 103751: contig of 4375 bp in length
103752 103851: gap of 100 bp
103852 108407: contig of 4556 bp in length
108408 108507: gap of 100 bp
108508 112468: contig of 3961 bp in length
112469 112568: gap of 100 bp
112569 116411: contig of 3843 bp in length
116412 116511: gap of 100 bp
116512 120690: contig of 4179 bp in length
120691 120790: gap of 100 bp
120791 125789: contig of 4999 bp in length
125790 125889: gap of 100 bp
125890 129429: contig of 3540 bp in length
129430 129529: gap of 100 bp
129530 132875: contig of 3346 bp in length
132876 132975: gap of 100 bp
132976 136718: contig of 3743 bp in length
136719 136818: gap of 100 bp
136819 140204: contig of 3386 bp in length
140205 140304: gap of 100 bp
140305 143212: contig of 2908 bp in length
143213 143312: gap of 100 bp
143313 146311: contig of 2999 bp in length
146312 146411: gap of 100 bp
146412 148694: contig of 2283 bp in length
148695 148794: gap of 100 bp
148795 151551: contig of 2757 bp in length
151552 151651: gap of 100 bp
151652 153471: contig of 1820 bp in length

Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,

acbrufe/s./ meadowe/s./ tttttt/s./ ttavels/m./ ttglttt/s./

Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 162162 bases at least Q40
 Consensus quality: 175389 bases at least Q30
 Consensus quality: 181326 bases at least Q20
 Insert size: 184151; sum-of-contigs
 Quality coverage: 4.54x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 42519 contig of 42519 bp in length
42620 59852 contig of 17233 bp in length
59953 74496 contig of 14544 bp in length
74597 87440 contig of 12844 bp in length
87541 99577 contig of 12037 bp in length
99678 110659 contig of 10982 bp in length
110760 115413 contig of 4654 bp in length
115514 124990 contig of 9477 bp in length
125091 135744 contig of 10654 bp in length
135845 145972 contig of 10128 bp in length
146073 153170 contig of 7098 bp in length
153271 156398 contig of 3128 bp in length
156499 161724 contig of 5226 bp in length
161825 165688 contig of 3964 bp in length
165789 169698 contig of 3910 bp in length
169799 173554 contig of 3756 bp in length
173655 175764 contig of 2110 bp in length
175865 177303 contig of 1439 bp in length
179198 181636 contig of 2439 bp in length
181737 183644 contig of 1908 bp in length
183745 184761 contig of 1017 bp in length
184862 186351 contig of 1490 bp in length

```

Sequence updated (26-May-2000).

NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 42519: contig of 42519 bp in length
42620 42619: gap of 100 bp
42620 59852: contig of 17233 bp in length
59853 59952: gap of 100 bp
59953 74496: contig of 14544 bp in length
74497 74596: gap of 100 bp
74597 87440: contig of 12844 bp in length
87441 87540: gap of 100 bp
87541 99577: contig of 12037 bp in length
99578 99677: gap of 100 bp
99678 110659: contig of 10982 bp in length
110660 110759: gap of 100 bp
110760 115413: contig of 4654 bp in length
115414 115513: gap of 100 bp
115514 124990: contig of 9477 bp in length
124991 125090: gap of 100 bp
125091 135744: contig of 10654 bp in length
135745 135844: gap of 100 bp
135845 145972: contig of 10128 bp in length
145973 146072: gap of 100 bp
146073 153170: contig of 7098 bp in length
153171 153270: gap of 100 bp
153271 156398: contig of 3128 bp in length
156399 156498: gap of 100 bp
156499 161724: contig of 5226 bp in length
161725 161824: gap of 100 bp

```

```

* 161825 165688: contig of 3864 bp in length
* 165689 165788: gap of 100 bp
* 165789 169698: contig of 3910 bp in length
* 169699 169798: gap of 100 bp
* 169799 173554: contig of 3756 bp in length
* 173555 173654: gap of 100 bp
* 173655 175764: contig of 2110 bp in length
* 175765 175864: gap of 100 bp
* 175865 177303: contig of 1439 bp in length
* 177304 177403: gap of 100 bp
* 177404 179097: contig of 1694 bp in length
* 179098 179197: gap of 100 bp
* 179198 181636: contig of 2439 bp in length
* 181637 181736: gap of 100 bp
* 181737 183644: contig of 1908 bp in length
* 183645 183744: gap of 100 bp
* 183745 184761: contig of 1017 bp in length
* 184762 184861: gap of 100 bp
* 184862 186351: contig of 1490 bp in length.

```

FEATURES

Source

```

1. 186351
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="18"
   /map="18q22"
   /clone="RP11-879N20"
1. 42519
   /note="assembly_fragment"
42620. 59852
   /note="assembly_fragment"
59953. 74496
   /note="assembly_fragment"
74597. 87440
   /note="assembly_fragment"
87541. 99577
   /note="assembly_fragment"
99678. 110659
   /note="assembly_fragment"
110760. 115413
   /note="assembly_fragment clone_end:SP6 vector_side:right"
115514. 124990
   /note="assembly_fragment"
125091. 135744
   /note="assembly_fragment"
135845. 145972
   /note="assembly_fragment"
146073. 153170
   /note="assembly_fragment"
153271. 156398
   /note="assembly_fragment clone_end:T7 vector_side:left"
156499. 161724
   /note="assembly_fragment"
161825. 165688
   /note="assembly_fragment"
165789. 169698
   /note="assembly_fragment"
169799. 173554
   /note="assembly_fragment"
173655. 175764
   /note="assembly_fragment"
175865. 177303
   /note="assembly_fragment"
177404. 179097
   /note="assembly_fragment"
179198. 181636
   /note="assembly_fragment"
181737. 183644
   /note="assembly_fragment"
183745. 184761
   /note="assembly_fragment"
184862. 186351
   /note="assembly_fragment"

```

misc_feature

```

misc_feature 42620. 59852
misc_feature 59953. 74496
misc_feature 74597. 87440
misc_feature 87541. 99577
misc_feature 99678. 110659
misc_feature 110760. 115413
misc_feature 115514. 124990
misc_feature 125091. 135744
misc_feature 135845. 145972
misc_feature 146073. 153170
misc_feature 153271. 156398
misc_feature 156499. 161724
misc_feature 161825. 165688
misc_feature 165789. 169698
misc_feature 169799. 173554
misc_feature 173655. 175764
misc_feature 175865. 177303
misc_feature 177404. 179097
misc_feature 179198. 181636
misc_feature 181737. 183644
misc_feature 183745. 184761
misc_feature 184862. 186351
misc_feature 186351. 186351

```

BASE COUNT 56377 a 36459 c 35952 g 55363 t 2200 others

STS

```

/repeat_region /standard_name="4let7"
/db_xref="dbSTS:G31407"
45099..45386
/rpt_family="Alu"
/misc_feature complement(48265..48384)
/note="GRAIL 2 excellent exon, frame 0"
/misc_feature complement(4958..50004)
/note="GRAIL 2 excellent exon, frame 0"
/repeat_region complement(50482..50554)
/rpt_family="Alu"
/repeat_region complement(50903..50972)
/rpt_family="Alu"
/repeat_region complement(51301..51570)
/rpt_family="Alu"
/repeat_region 53075..53337
/rpt_family="Alu"
/misc_feature complement(59112..59259)
/note="GRAIL 2 excellent exon, frame 0"
/repeat_region 61720..61983
/rpt_family="Alu"
/repeat_region 61998..62033
/note="(AATA)9"
/rpt_type=tandem
/misc_feature /rpt_unit=AATA
66046..66147
/repeat_region /note="(ATCT)10"
68813..68852
/rpt_type=tandem
/repeat_region /rpt_unit=ATCT
68852..68892
/rpt_family="MIR2"
/misc_feature 72177..72322
/standard_name="CDC1e3"
/note="99% identity B07638 (exon trapped product)"
/repeat_region 72614..72649
/note="(AC)18"
/rpt_type=tandem
/repeat_region /rpt_unit=AC
73695..73695)
/rpt_family="Alu"
/misc_feature complement(73801..73925)
/note="8% identity EST oe52h01.s1"
/db_xref="dbEST:AA862321"
73944..74270
/standard_name="LTR7"
/repeat_region 76779..77132
/rpt_family="THE1"
/repeat_region 77616..77670
/rpt_family="MADE1"
/repeat_region complement(80901..81095)
/rpt_family="Alu"
/misc_feature 81076..81185
/note="GRAIL 2 excellent exon, frame 2"
/repeat_region complement(81404..81485)
/rpt_family="MIR"
/repeat_region complement(82147..82531)
/rpt_family="THE1"
/misc_feature complement(82908..83342)
/note="99% identity EST Y982c08.r1"
/db_xref="dbEST:R53285"
83752..84040
/rpt_family="Alu"
/repeat_region 84023..84044
/note="(A)22"
/rpt_type=tandem
/rpt_unit=A

```

Query Match 80.9% Score 17.8; DB 9; Length 95448;
 Best Local Similarity 90.5%; Pred. NO. 99;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCAGACGGTCTGAGTCATCT 22

```

Db 44786 TCAGACGGTCTGAGTCATCT 44766
|||||
CNS08C9K 124092 bp DNA linear HTG 04-JUN-2002
Oryza sativa chromosome 12 clone OSUNB0090H23, *** SEQUENCING IN
PROGRESS ***, 11 ordered pieces.
ACCESSION AL732645
VERSION AL732645.1 GI:21326737
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 124092)
AUTHORS Cholsne,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P.,
Segurens,B., Pelletier,E., Scarpelli,C., Salanoubat,M.,
Weissenbach,J. and Quetier,F.
Oryza sativa chromosome 12 sequencing
Unpublished
2 (bases 1 to 124092)
Genoscope.
Direct Submission
Submitted (03-JUN-2002) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence.
Work on the sequence is in progress and the release of this data is
based on the understanding that the sequence may change as work
continue. The sequence may be contaminated with foreign sequence
from E.coli, yeast, vector, phage, etc.
Contigs composition :
1707 bp contig from 1 to 1707
13309 bp contig from 1808 to 15116
6358 bp contig from 15217 to 21574
22606 bp contig from 21675 to 44280
9040 bp contig from 44381 to 53420
20348 bp contig from 53521 to 73868
9188 bp contig from 73969 to 83156
19261 bp contig from 83257 to 102517
6512 bp contig from 102618 to 109129
8922 bp contig from 109230 to 118151
5841 bp contig from 118252 to 124092.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 1708 1807: gap of 100 bp
* 1808 15116: contig of 13309 bp in length
* 15117 15216: gap of 100 bp
* 15217 21574: contig of 6358 bp in length
* 21575 21674: gap of 100 bp
* 21675 44280: contig of 22606 bp in length
* 44281 44380: gap of 100 bp
* 44381 53420: contig of 9040 bp in length
* 53421 53520: gap of 100 bp
* 53521 73868: contig of 20348 bp in length
* 73869 73968: gap of 100 bp
* 73969 83156: contig of 9188 bp in length
* 83157 83256: gap of 100 bp
* 83257 102517: contig of 19261 bp in length
* 102518 102617: gap of 100 bp
* 102618 109129: contig of 6512 bp in length

```

* 109130 109229: gap of 100 bp
 * 109230 118151: contig of 8922 bp in length
 * 118152 118251: gap of 100 bp
 * 118252 124092: contig of 5841 bp in length.

FEATURES

Location/Qualifiers
 1..124092
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="12"
 /clone="OSJNB0090H23"
 /clone.lib="OSJNBb"

BASE COUNT 33707 a 27560 c 27075 g 34750 t 1000 others
 ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 124092;
 Best Local Similarity 90.5%; Pred. No. 1e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCGACGGTCTGAGTCATC 21
 |||||

Db 90490 ACCGACGGTCTGAGTCATC 90510

RESULT 10

AL606903 175238 bp DNA linear ROD 20-FEB-2002
 LOCUS Mouse DNA sequence from clone Rp23-27813 on chromosome 4, complete
 DEFINITION sequence.

ACCESSION AL606903.8 GI:18873508

VERSION HTG.

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
 Garner.P.

REFERENCE 1 (bases 1 to 175238)

TITLE Direct Submission

JOURNAL

Submitted (20-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Feb 22, 2002 this sequence version replaced gi:17381401.
 During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em:, EMBL; Sw:,
 SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
 database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep Rp23-27813 is

from the RCI-23 Mouse PAC library

constructed by the group of Pieter de Jong.

For further details see http://www.chori.org/bacpac/home.htm

VECTOR: pBACe3.6

This sequence is the entire insert of clone Rp23-27813.

Location/Qualifiers

1..175238
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="4"
 /clone="Rp23-27813"

FEATURES

source

1 1667: contig of 1667 bp in length
 * 1668 1767: gap of unknown length
 * 1768 3037: contig of 1270 bp in length
 * 3038 3137: gap of unknown length
 * 3138 4370: contig of 1233 bp in length
 * 4371 4471: gap of unknown length
 * 4471 6092: contig of 1622 bp in length
 * 6093 6192: gap of unknown length

BASE COUNT 45250 a 39075 c 40628 g 50285 t
 ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 175238;
 Best Local Similarity 90.5%; Pred. No. 1.1e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCGACGGTCTGAGTCATC 22
 |||||

Db 11296 TCGATCGTCTGAGTCATC 11316

RESULT 11

LOCUS

AC084869 176641 bp DNA linear HTG 11-MAR-2001
 DEFINITION Homo sapiens chromosome RCI-11 clone RP11-508E4, WORKING DRAFT
 SEQUENCE, 44 unordered pieces.

ACCESSION AC084869

VERSION AC084869.2 GI:11465172

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 176641)

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 176641)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (24-NOV-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT On Nov 30, 2000 this sequence version replaced gi:11323446.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----

Center project name: H.NH0508E04

----- Summary Statistics -----

Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%

Chemistry: Dye-terminator Big Dye; 0% of reads

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 142830 bases at least Q40

Consensus quality: 154079 bases at least Q30

Consensus quality: 160559 bases at least Q20

Insert size: 183000; agarose-fp

Insert size: 172341; sum-of-contigs

Quality coverage: 2.73 in Q20 bases; agarose-fp

Quality coverage: 3.02 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 44 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 6193 7578: contig of 1386 bp in length
* 7579 7678: gap of unknown length
* 7580 9287: contig of 1609 bp in length
* 9288 9387: gap of unknown length
* 9388 11041: contig of 1654 bp in length
* 11042 11141: gap of unknown length
* 11142 12827: contig of 1686 bp in length
* 12828 12927: gap of unknown length
* 12928 15022: contig of 2095 bp in length
* 15023 15122: gap of unknown length
* 15123 16154: contig of 1032 bp in length
* 16155 16255: gap of unknown length
* 16256 17925: contig of 1670 bp in length
* 17926 18025: gap of unknown length
* 18026 20170: contig of 2146 bp in length
* 20171 20270: gap of unknown length
* 20271 21181: contig of 2911 bp in length
* 21182 22281: gap of unknown length
* 22282 23440: contig of 2059 bp in length
* 23441 25441: gap of unknown length
* 25442 28590: contig of 3150 bp in length
* 28591 28690: gap of unknown length
* 28691 31015: contig of 2325 bp in length
* 31016 31115: gap of unknown length
* 31116 33667: contig of 2752 bp in length
* 33668 37010: contig of 3043 bp in length
* 37011 37110: gap of unknown length
* 37111 39680: contig of 2570 bp in length
* 39681 39780: gap of unknown length
* 39781 42432: contig of 2652 bp in length
* 42433 42533: gap of unknown length
* 42534 45045: contig of 2512 bp in length
* 45046 45144: gap of unknown length
* 45145 48026: contig of 2882 bp in length
* 48027 48126: gap of unknown length
* 48127 51134: contig of 3008 bp in length
* 51135 51235: gap of unknown length
* 51236 53679: contig of 2445 bp in length
* 53680 53779: gap of unknown length
* 53780 56546: contig of 2767 bp in length
* 56547 56646: gap of unknown length
* 56647 59359: contig of 2713 bp in length
* 59360 59459: gap of unknown length
* 59460 62921: contig of 3462 bp in length
* 62922 63021: gap of unknown length
* 63022 65458: contig of 2437 bp in length
* 65459 65559: gap of unknown length
* 65560 68534: contig of 3976 bp in length
* 68535 69634: gap of unknown length
* 69635 73068: contig of 3434 bp in length
* 73069 73168: gap of unknown length
* 73169 78010: contig of 4842 bp in length
* 78011 78110: gap of unknown length
* 78111 83482: contig of 5372 bp in length
* 83483 83582: gap of unknown length
* 83583 88800: contig of 5218 bp in length
* 88801 89900: gap of unknown length
* 89901 97207: contig of 8307 bp in length
* 97208 97307: gap of unknown length
* 97308 104462: contig of 7155 bp in length
* 104463 104562: gap of unknown length
* 104563 111426: contig of 6864 bp in length
* 111427 111526: gap of unknown length
* 111527 118832: contig of 7306 bp in length
* 118833 118932: gap of unknown length
* 118933 125694: contig of 6762 bp in length
* 125695 125794: gap of unknown length
* 125795 133108: contig of 7314 bp in length
* 133109 133208: gap of unknown length
* 133209 141040: contig of 7832 bp in length
* 141041 141140: gap of unknown length
* 141141 150417: contig of 9277 bp in length

* 150418 150517: gap of unknown length
* 150518 158426: contig of 7909 bp in length
* 158427 158526: gap of unknown length
* 158527 167566: contig of 9040 bp in length
* 167567 167666: gap of unknown length
* 167667 176641: contig of 8975 bp in length.
FEATURES
 source
 1..176641
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="RPCL11"
 /clone="RP11-508E4"
 1..1667
 /note="assembly_name:Contig15"
 1768..3037
 /note="assembly_name:Contig16"
 3138..4370
 /note="assembly_name:Contig18"
 4471..6092
 /note="assembly_name:Contig19"
 6193..7578
 /note="assembly_name:Contig20"
 7679..9287
 /note="assembly_name:Contig21"
 9388..11041
 /note="assembly_name:Contig22"
 11142..12827
 /note="assembly_name:Contig23"
 12928..15022
 /note="assembly_name:Contig24"
 15123..16154
 /note="assembly_name:Contig25"
 16255..17924
 /note="assembly_name:Contig26"
 18025..20170
 /note="assembly_name:Contig27"
 20271..23181
 /note="assembly_name:Contig28"
 23282..25340
 /note="assembly_name:Contig29"
 25441..28590
 /note="assembly_name:Contig30"
 28691..31015
 /note="assembly_name:Contig31"
 31116..33867
 /note="assembly_name:Contig32"
 33968..37010
 /note="assembly_name:Contig33"
 37111..39680
 /note="assembly_name:Contig34"
 39781..42432
 /note="assembly_name:Contig35"
 42533..45044
 /note="assembly_name:Contig36"
 45145..48026
 /note="assembly_name:Contig37"
 48127..51134
 /note="assembly_name:Contig38"
 51235..53679
 /note="assembly_name:Contig39"
 53780..56546
 /note="assembly_name:Contig40"
 56647..59359
 /note="assembly_name:Contig41"
 59460..62921
 /note="assembly_name:Contig42"
 63022..65458
 /note="assembly_name:Contig43"
 65559..69534

Query Match 80.9%; Score 17.8; DB 2; Length 176641;
Best Local Similarity 90.5%; Pred.No.1.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

44171 44270: gap of unknown length
 44271 46570: contig of 2300 bp in length
 46571 46670: gap of unknown length
 46671 49909: contig of 3239 bp in length
 49910 50009: gap of unknown length
 50010 53419: contig of 3410 bp in length
 53420 53519: gap of unknown length
 53520 56387: contig of 2868 bp in length
 56388 60933: contig of 4445 bp in length
 60933 61033: gap of unknown length
 61033 63607: contig of 2575 bp in length
 63607 63707: gap of unknown length
 63707 69387: contig of 5680 bp in length
 69387 69487: gap of unknown length
 69487 71963: contig of 2476 bp in length
 71963 72063: gap of unknown length
 72063 75957: contig of 3894 bp in length
 75957 76057: gap of unknown length
 76057 79570: contig of 3512 bp in length
 79570 85878: gap of unknown length
 85878 90719: contig of 6208 bp in length
 90719 90818: gap of unknown length
 90818 94719: contig of 3900 bp in length
 94719 94818: gap of unknown length
 94818 100271: contig of 5453 bp in length
 100271 100372: gap of unknown length
 100372 106350: contig of 5879 bp in length
 106350 106351: gap of unknown length
 106351 112115: contig of 5765 bp in length
 112115 112215: gap of unknown length
 112215 119503: contig of 7288 bp in length
 119503 119504: gap of unknown length
 119504 125419: contig of 5816 bp in length
 125419 125519: gap of unknown length
 125519 133159: contig of 7640 bp in length
 133159 133259: gap of unknown length
 133259 140941: contig of 7682 bp in length
 140941 141041: gap of unknown length
 141041 147717: contig of 6676 bp in length
 147717 147817: gap of unknown length
 147817 157811: contig of 9994 bp in length
 157811 157911: gap of unknown length
 157911 169168: contig of 11257 bp in length
 169168 169268: gap of unknown length
 169268 178103: contig of 8835 bp in length.
 Location/Qualifiers
 1. .178103
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /clone="CH230-246K21"

BASE COUNT 47068 a 39502 c 39309 g 47550 t 4674 others
 ORIGIN
 Query Match 80.9%; Score 17.8; DB 2; Length 178103;
 Best Local Similarity 90.5%; Pred. No. 1.1e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCGAAGGTTCTGAGTCATC 21
 ||||| ||||| ||||| |||||
 Db 161889 ATCGAAGGTTCTGAGTCATC 161869

RESULT 13
 AL606971 228664 bp DNA linear HTG 13-AUG-2002
 LOCUS
 DEFINITION Mus musculus chromosome 4 clone RP23-123120, *** SEQUENCING IN
 PROGRESS ***, 2 unordered pieces.
 ACCESSION AL606971
 VERSION AL606971.9 GI:21689948
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 228664)
 AUTHORS Ramsay H.
 TITLE Direct Submission
 JOURNAL Submitted (08-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jul 3, 2002 this sequence version replaced gi:21261813.
 COMMENT
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BM123120
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator; 2% of reads
 Chemistry: Dye-terminator Big Dye; 97% of reads
 Consensus quality: 228120 bases at least Q40
 Consensus quality: 228283 bases at least Q30
 Consensus quality: 228383 bases at least Q20
 Insert size: 228564; sum-of-contigs
 Insert size: 203349; 5.9% error; agarose-fp
 Quality coverage: 11.87x in Q20 bases; sum-of-contigs Quality
 coverage: 13.78x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 138728: contig of 138728 bp in length
 * 138729 138828: gap of 100 bp
 * 138829 228664: contig of 89836 bp in length.
 FEATURES
 source
 1. .228664
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="4"
 /clone="RP23-123120"
 /clone_lib="RPCI-23"
 misc_feature
 1. .138728
 /note="assembly_fragment:02064"
 misc_feature
 138829. .228664
 /note="assembly_fragment:07697"
 BASE COUNT 56606 a 56012 c 56424 g 59522 t 100 others
 ORIGIN
 Query Match 80.9%; Score 17.8; DB 2; Length 228664;
 Best Local Similarity 90.5%; Pred. No. 1.1e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 TCGAAGGTTCTGAGTCATCT 22
 ||||| ||||| ||||| |||||
 Db 51570 TCGAAGGTTCTGAGTCATCT 51590
 RESULT 14
 AC010433/c 258929 bp DNA linear PRI 10-JUL-2002
 LOCUS
 DEFINITION Homo sapiens chromosome 5 clone CTD-2202L20, complete sequence.
 ACCESSION AC010433
 VERSION AC010433.9 GI:21717106
 KEYWORDS HTG.
 SOURCE human.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 19:55:35 ; Search time 10.4381 Seconds
(without alignments)
4746.432 Million cell updates/sec

Title: US-09-581-500B-13

Perfect score: 22

Sequence: 1 atcgaaaggttgtgagtcattc 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID32/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID32/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match %	Length	DB ID	Description
1	22	100.0	22	20	AA388557	Human chromosome 1
2	22	100.0	656	20	AA388553	Human chromosome 1
3	17.2	78.2	486	20	AAV87361	EST clone BV275.
4	17.2	78.2	1164	21	AA418397	Arabidopsis thalia
5	17.2	78.2	1522	21	AAC77253	Human ORFX ORF2808
6	17.2	78.2	1689	14	AAQ52498	Helminth aminopept
7	17.2	78.2	3006	14	AAQ52491	Helminth aminopept
8	17.2	78.2	3084	14	AAQ52500	Helminth aminopept
9	17.2	78.2	10347	22	AAF28060	Human 07CG27 gene

10 17.2 78.2 13202 24 AB333484 Human immune syste
c 11 17.2 78.2 13202 24 AB333485 Human immune syste
c 12 16.2 73.6 388 21 AAC27333 Human secreted pro
c 13 16.2 73.6 458 21 AAC40995 Zea mays DNA fragm
14 16.2 73.6 470 23 ABV13083 Human prostate exp
15 16.2 73.6 911 21 AAC43639 Zea mays DNA fragm
c 16 16.2 73.6 1030 21 AAC34348 Arabidopsis thalia
c 17 16.2 73.6 1394 9 AAN82030 Synthetic Hydroxym
c 18 16.2 73.6 1394 14 AAQ46621 Encodes human 3-hy
c 19 16.2 73.6 15320 23 ABL07834 Drosophila melanog
c 20 15.8 71.8 1203 24 ABN68044 Streptococcus poly
c 21 15.8 71.8 1293 24 ABK72339 Bacillus lichenifo
c 22 15.8 71.8 1527 21 AAC43261 Arabidopsis thalia
c 23 15.8 71.8 1662 24 ABL41878 Nucleotide sequenc
c 24 15.6 70.9 292 24 ABL71087 Corn tassal-derive
c 25 15.6 70.9 398 21 AAA79317 Eucalyptus grandis
c 26 15.6 70.9 461 21 AAA79313 Eucalyptus grandis
c 27 15.6 70.9 1266 21 AAC51603 Arabidopsis thalia
c 28 15.6 70.9 1499 21 AAC39132 Arabidopsis thalia
c 29 15.6 70.9 1619 17 AAT18796 Human immunophilin
c 30 15.6 70.9 2447 15 AAQ56702 Partial sequence o
c 31 15.6 70.9 2447 22 AAD11038 Human kappa opioid
c 32 15.6 70.9 2719 23 ABL28558 Drosophila melanog
c 33 15.6 70.9 3109 21 AAC78150 Human cancer assoc
c 34 15.6 70.9 3528 22 AAS23393 Candida albicans e
c 35 15.6 70.9 147419 24 ABK83574 Human cDNA differe
c 36 15.6 70.9 325791 22 AAS43104 Human Oestrogen re
c 37 15.6 70.9 1830121 17 AAT42063 Haemophilus influe
c 38 15.6 70.9 2944528 24 ABA03041 Listeria monocytog
c 39 15.4 70.0 723 23 AAS66765 DNA encoding novel
c 40 15.4 70.0 1117 21 AAC35713 Arabidopsis thalia
c 41 15.4 70.0 1117 21 AAC51417 Arabidopsis thalia
c 42 15.4 70.0 1650 22 AAI59082 Human polynucleoti
c 43 15.4 70.0 1711 19 AAV33136 Plasmodium berghei
c 44 15.4 70.0 1735 22 AAK52200 Human polynucleoti
c 45 15.4 70.0 2168 21 AAC59479 Human secreted pro

ALIGNMENTS

RESULT 1

AA388557

ID AAX388557 standard; DNA; 22 BP.

XX AC AAX88557;

XX XX

DT 10-SEP-1999 (first entry)

XX Human chromosome 18q YAC clone amplification primer.

DE Human chromosome 18q; mood disorder; polymorphic marker; detection;

DE Human chromosome 18q; mood disorder; polymorphic marker; detection;

KW identification; trinucleotide repeat expansion; schizophrenia;

KW anxiety disorder; adjustment disorder; personality disorder;

KW nucleotide triplet repeat; ss.

XX Synthetic.

OS Homo sapiens.

XX WO9932643-A2.

XX 01-JUL-1999.

XX 17-DEC-1998; 98WO-EP08543.

XX 18-DEC-1997; 97GB-0026804.

XX (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.

XX Del-Favero J, Raeymaekers P, Van Broeckhoven C;

XX WPI; 1999-418934/35.

XX

PT Detecting nucleotide triplet repeats in human chromosome 18q

PS Claim 29; Fig 15b; 87pp; English.

XX The present invention describes detecting nucleotide triplet repeats in
 CC a region of human chromosome 18q disposed between polymorphic markers
 CC D18S68 and D18S979 to identify a human gene associated with a mood
 CC disorder or related disorder. AAX88542 to AAX88705 represents human
 CC chromosome 18q YAC clones and primers corresponding to them, used in the
 CC exemplification of the present invention. YAC clones comprising a
 CC portion of the region of human chromosome 18q between markers D18S68 and
 CC D18S979 are used to identify at least one human gene associated with a
 CC mood disorder or related disorder. The mood disorder or related
 CC disorder, is chosen from the Diagnostic and Statistical Manual of Mental
 CC Disorders, version 4 (DSM-IV) taxonomy. This includes mood disorders
 CC (296.XX, 300.4, 311, 301, 13, 295.70), schizophrenia and related
 CC disorders (295, 297.1, 298.9, 297.3, 298.9), anxiety disorders (300.XX,
 CC 309.81, 308.3), adjustment disorders (309.XX) and personality disorders
 CC (codes 301.XX). Probes derived from genes associated with the mood
 CC disorder or related disorder can be used to detect pathological
 CC mutations or genetic variations in patients. The methods, probes and
 CC antibodies can be used to determine the susceptibility of an individual
 CC to a mood disorder or related disorder. The nucleic acids and proteins
 CC of the human gene can be used to treat mood disorders and related
 CC disorders.

XX Sequence 22 BP; 5 A; 5 C; 5 G; 7 T; 0 other;

Query Match 100.0%; Score 22; DB 20; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.074;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTCTCGAGTCATCT 22

DB 1 ATCGACGGTCTCGAGTCATCT 22

RESULT 2

AA88553

ID AAX88553 standard; DNA; 656 BP.

XX

AC AAX88553;

XX 10-SEP-1999 (first entry)

DE Human chromosome 18q YAC clone nucleotide sequence #12.

XX Human chromosome 18q; mood disorder; polymorphic marker; detection;
 KW identification; trinucleotide repeat expansion; schizophrenia;
 KW anxiety disorder; adjustment disorder; personality disorder;
 KW nucleotide triplet repeat; ss.

OS Homo sapiens.

OS Synthetic.

XX WO9932643-A2.

XX 01-JUL-1999.

XX 17-DEC-1998; 98WO-BP08543.

XX 18-DEC-1997; 97GB-0026804.

XX (VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
 XX Del-Favero J, Raeymaekers P, Van Broeckhoven C;
 XX WPI; 1999-418934/35.

XX Detecting nucleotide triplet repeats in human chromosome 18q
 PT Claim 21; Fig 15a; 87pp; English.

PS

XX

CC The present invention describes detecting nucleotide triplet repeats in
 CC a region of human chromosome 18q disposed between polymorphic markers
 CC D18S68 and D18S979 to identify a human gene associated with a mood
 CC disorder or related disorder. AAX88542 to AAX88705 represents human
 CC chromosome 18q YAC clones and primers corresponding to them, used in the
 CC exemplification of the present invention. YAC clones comprising a
 CC portion of the region of human chromosome 18q between markers D18S68 and
 CC D18S979 are used to identify at least one human gene associated with a
 CC mood disorder or related disorder. The mood disorder or related
 CC disorder, is chosen from the Diagnostic and Statistical Manual of Mental
 CC Disorders, version 4 (DSM-IV) taxonomy. This includes mood disorders
 CC (296.XX, 300.4, 311, 301, 13, 295.70), schizophrenia and related
 CC disorders (295, 297.1, 298.9, 297.3, 298.9), anxiety disorders (300.XX,
 CC 309.81, 308.3), adjustment disorders (309.XX) and personality disorders
 CC (codes 301.XX). Probes derived from genes associated with the mood
 CC disorder or related disorder can be used to detect pathological
 CC mutations or genetic variations in patients. The methods, probes and
 CC antibodies can be used to determine the susceptibility of an individual
 CC to a mood disorder or related disorder. The nucleic acids and proteins
 CC of the human gene can be used to treat mood disorders and related
 CC disorders.

XX Sequence 656 BP; 183 A; 165 C; 208 G; 96 T; 4 other;

Query Match 100.0%; Score 22; DB 20; Length 656;

Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTCTCGAGTCATCT 22

DB 278 ATCGACGGTCTCGAGTCATCT 299

RESULT 3

AA87361

ID AAV87361 standard; cDNA; 486 BP.

XX

AC AAV87361;

XX 27-APR-1999 (first entry)

XX EST clone BV275.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibitor; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

XX WO9845435-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-US06954.

XX 10-APR-1997; 97US-0835913.

XX (GEMY) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX Racle IA, Spaulding V, Treacy M;

XX WPI; 1999-070076/06.

XX New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 XX Claim 1; Page 540; 633pp; English.

XX This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are

PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 08-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.
 PR 29-OCT-1999; 99US-0162142.
 Query Match 78.2%; Score 17.2; DB 21; Length 1164;
 Best Local Similarity 86.4%; Pred. No. 39;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 ATCGAACGGTTCGATCATCT 22
 Db 201 ACCGAATCGTTCGATCATCT 222
 RESULT 5
 AAC77253
 ID AAC77253 standard; cDNA; 1522 BP.
 AC AAC77253;
 XX
 XX 08-FEB-2001 (first entry)
 DT Human ORFX ORF2808 polynucleotide sequence SEQ ID NO:5615.
 DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; nontropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX Homo sapiens.
 OS
 XX WC200058473-A2.
 PN
 XX 05-OCT-2000.
 PD
 XX 31-MAR-2000; 2000WO-US08621.
 PF
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 PA
 XX Shimkets RA, Leach M;
 PI
 XX WPI; 2000-602362/57.
 DR P-FSDB; AAB43044.
 DR
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 4799-4800; 5507pp; English.
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; nontropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 1522 BP; 537 A; 352 C; 324 G; 309 T; 0 other;

Query Match 78.2%; Score 17.2; DB 21; Length 1522;
 Best Local Similarity 86.4%; Pred. No. 40;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCGTGACTCATCT 22
 | ||||| ||||| ||||| |||||
 Db 369 ACCGACACAGTTCGTGACTCATTT 390

RESULT 6
 AAQ52498/c
 ID AAQ52498 standard; cDNA; 1689 BP.

XX AAQ52498;
 XX
 DT 31-MAY-1994 (first entry)
 XX

DE Helminth aminopeptidase H11-1 clone AustBI.

XX Helminth; aminopeptidase; H11-3; H11-2; H11-1; integral membrane;
 KW alpha-amino acyl peptide hydrolase (microsomal); multigene family;
 KW antigen; vaccine; parasite; human; virus; microbe; H110D; ss.
 XX

OS Haemonchus contortus.

XX WO9323542-A.

PN 25-NOV-1993.

XX 07-MAY-1993; 93WO-GB00943.

XX 08-MAY-1992; 92GB-0009993.

XX (AGRI-) AGRIC & FOOD RES COUNCIL.

XX Graham M, Knox DP, Munn EA, Newton SE, Oliver JJ;

PI Smith TS;

XX WPI; 1993-386574/48.

XX New DNA encoding amino-peptidase from Haemonchus contortus and
 PT derived peptide(s) - useful in protective vaccines, for new
 PT vectors, transformed cells and oligosaccharide, for incorporation
 PT in virus or microbe
 XX

XX Claim 1; Fig 5; 137pp; English.

XX The sequences given in AAQ52498-99 represent cDNA clones of the helminth
 CC aminopeptidase gene H11-1. H11-1 encodes H110D which is a protein
 CC doublet which shows homology to a family of integral membrane amino-
 CC peptidases. The differences between these coding sequences can be
 CC attributed to different mRNAs of the multigene family, and also to
 CC different variants of the H110D-encoding sequence being present at
 CC different stages of the life cycle, or in strains differing in

CC geographical origin. Antigenic fragments of the aminopeptidases
 CC encoded by these sequences may be used in vaccines to stimulate
 CC immune response against helminth parasites in humans or other animals.
 CC These DNA sequences may be incorporated into a virus or microbe and
 CC used in a similar manner.

XX SQ Sequence 1689 BP; 515 A; 330 C; 419 G; 425 T; 0 other;

Query Match 78.2%; Score 17.2; DB 14; Length 1689;
 Best Local Similarity 86.4%; Pred. No. 41;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCGTGACTCATCT 22
 | ||||| ||||| ||||| |||||
 Db 505 ATCGAACGGTTCGTGACTCATCT 484

RESULT 7

AAQ52491/c
 ID AAQ52491 standard; DNA; 3006 BP.

XX AAQ52491;

XX DT 31-MAY-1994 (first entry)

XX DE Helminth aminopeptidase H11-1 gene.

XX Helminth; aminopeptidase; H11-3; H11-2; H11-1; integral membrane;
 KW alpha-amino acyl peptide hydrolase (microsomal); multigene family;
 KW antigen; vaccine; parasite; human; virus; microbe; ss.
 XX

OS Haemonchus contortus.

XX Key Location/Qualifiers
 FH CDS 33..3006
 FT /*tag= a
 FT /product= Aminopeptidase H11-3

XX WO9323542-A.

XX 25-NOV-1993.

XX 07-MAY-1993; 93WO-GB00943.

XX 08-MAY-1992; 92GB-0009993.

XX (AGRI-) AGRIC & FOOD RES COUNCIL.

XX Graham M, Knox DP, Munn EA, Newton SE, Oliver JJ;

PI Smith TS;

XX WPI; 1993-386574/48.

XX P-PSDB; AAR51282.

XX New DNA encoding amino-peptidase from Haemonchus contortus and
 PT derived peptide(s) - useful in protective vaccines, for new
 PT vectors, transformed cells and oligosaccharide, for incorporation
 PT in virus or microbe
 XX

XX Claim 1; Fig 2; 137pp; English.

XX The sequences given in AAQ52489-91 represent the helminth amino-
 CC peptidase genes H11-3, -2 and -1 respectively. These sequences
 CC encode H110D which is a protein doublet which shows homology to a
 CC family of integral membrane aminopeptidases. The differences
 CC between these coding sequences can be attributed to different mRNAs
 CC of the multigene family, and also to different variants of the H110D-
 CC encoding sequence being present at different stages of the life
 CC cycle, or in strains differing in geographical origin. Antigenic
 CC fragments of the aminopeptidases encoded by these sequences may be
 CC used in vaccines to stimulate immune response against helminth
 CC parasites in humans or other animals. These DNA sequences may be
 CC incorporated into a virus or microbe and used in a similar manner.

XX SQ Sequence 3006 BP; 932 A; 600 C; 716 G; 758 T; 0 other;
 Query Match 78.2%; Score 17.2; DB 14; Length 3006;
 Best Local Similarity 86.4%; Pred. No. 44;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAAGGTTCTGAGTCATCT 22
 ||||| ||||| |||||
 Db 1429 ATCGAAGGTTCTGAGTCATCT 1408

RESULT 8
 AAF2500/c
 ID AAF2500 standard; DNA; 3084 BP.
 XX AC AAF2500;
 XX DT 31-MAY-1994 (first entry)
 XX DE Helminth aminopeptidase H11-1 gene.
 XX KW Helminth; aminopeptidase; H11-3; H11-2; H11-1; integral membrane;
 KW alpha-amino acyl peptide hydrolase (microsome); multigene family;
 KW antigen; vaccine; parasite; human; virus; microbe; ss.
 XX OS Haemonchus contortus.
 XX FH Key Location/Qualifiers
 FT CDS 23..2956
 FT /*tag= a
 FT /product= H110D variant
 XX WO9323542-A.
 XX PD 25-NOV-1993.
 XX PF 07-MAY-1993; 93WO-GB000943.
 XX PR 08-MAY-1992; 92GB-0009993.
 XX PA (AGRI-) AGRIC & FOOD RES COUNCIL.
 XX PI Graham M, Knox DP, Munn EA, Newton SE, Oliver JU;
 XX FT Smith TS;
 XX WPI; 1993-386574/48.
 XX DR P-PSDB; AAF51282.
 XX PS New DNA encoding amino-peptidase from Haemonchus contortus and
 PT derived peptide(s) - useful in protective vaccines, for new
 PT vectors, transformed cells and oligosaccharide, for incorporation
 PT in virus or microbe
 XX Claim 1; Page 75; 137pp; English.

The sequences given in AAF2500-02 represent the helminth amino-peptidase genes H11-1, -2 and -3 respectively. These sequences encode H110D which is a protein doublet which shows homology to a family of integral membrane aminopeptidases. The differences between these coding sequences can be attributed to different mRNAs of the multigene family, and also to different variants of the H110D-encoding sequence being present at different stages of the life cycle, or in strains differing in geographical origin. Antigenic fragments of the aminopeptidases encoded by these sequences may be used in vaccines to stimulate immune response against helminth parasites in humans or other animals. These DNA sequences may be incorporated into a virus or microbe and used in a similar manner.

SQ Sequence 3084 BP; 955 A; 610 C; 729 G; 790 T; 0 other;
 Query Match 78.2%; Score 17.2; DB 14; Length 3084;
 Best Local Similarity 86.4%; Pred. No. 44;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAAGGTTCTGAGTCATCT 22
 ||||| ||||| |||||
 Db 1429 ATCGAAGGTTCTGAGTCATCT 1408

RESULT 9
 AAF28060
 ID AAF28060 standard; CDNA; 10347 BP.
 XX AC AAF28060;
 XX DT 23-MAY-2001 (first entry)
 XX DE Human 07CG27 gene cDNA.
 XX KW Human; 07CG27 gene; chromosome 1; HPC1 region; prostate cancer;
 KW oncogene; ss.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT CDS 225..8684
 FT /*tag= a
 FT /product= "07CG27 protein"
 XX WO200116291-A2.
 XX PD 08-MAR-2001.
 XX PF 25-AUG-2000; 2000WO-US23391.
 XX PR 27-AUG-1999; 99US-0151049.
 XX PA (MYRI-) MYRIAD GENETICS INC.
 PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
 XX PT Tavtigian SV, Swedlund B, Simard J, Rommens JM;
 WPI; 2001-226682/23.
 XX DR P-PSDB; AAB35408.
 XX PT Novel human prostate cancer marker gene termed as 07CG27 gene, useful
 PT for screening mutations in the gene in diagnosis of a predisposition to
 PT cancer -
 XX Claim 2; Page 79-91; 99pp; English.

The present invention provides the protein and coding sequences of the human 07CG27 oncogene. This gene is found at the HPC1 region of chromosome 1. The sequences can be used in the diagnosis and identification of treatments for prostate cancer. The present sequence is the 07CG27 coding sequence.

SQ Sequence 10347 BP; 3325 A; 2451 C; 2229 G; 2342 T; 0 other;
 Query Match 78.2%; Score 17.2; DB 22; Length 10347;
 Best Local Similarity 86.4%; Pred. No. 52;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAAGGTTCTGAGTCATCT 22
 ||||| ||||| |||||
 Db 2567 ACCGAACAGTTCTGAGTCATTT 2588

RESULT 10
 ABL33484
 ID ABL33484 standard; DNA; 13202 BP.
 XX AC ABL33484;
 XX DT 26-MAR-2002 (first entry)

XX	Homo sapiens.
OS	WO200200928-A2.
XP	03-JAN-2002.
XX	
XX	02-JUL-2001; 2001WO-EP07537.
PF	
XX	
XX	30-JUN-2000; 2000DE-1032529.
PR	
XX	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
XX	
XX	WPI; 2002-130909/17.
DR	
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PPT	for diagnosis and treatment of diseases associated with abnormal
PTT	cytosine methylation -
XX	
XX	Claim 1; SEQ ID NO 1458; 32pp + Sequence Listing; German.
PS	
XX	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention.
XX	
XX	Sequence 13202 BP; 3984 A; 187 C; 2851 G; 6180 T; 0 other;
SQ	
	Query Match 78.2%; Score 17.2; DB 24; Length 13202;
	Best Local Similarity 86.4%; Pred. No. 54;
Matches	19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	1 ATCGAACGGTTCTGAGTCATCT 22
DB	8259 ATCGAAGATTCTAAATCATCT 8238
RESULT 12	
AAC27333/c	
ID AAC27333 standard; cDNA; 388 BP.	
XX AAC27333;	
XX	
DT 06-OCT-2000 (first entry)	
XX	
DE Human secreted protein 5' EST, SEQ ID NO: 31408.	
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;	
KW gene therapy; chromosome mapping; ss.	
XX	
OS Homo sapiens.	
XX	
PN EP1033401-A2.	
XX	
PD 06-SEP-2000.	
XX	
XX 21-FEB-2000; 2000EP-0200610.	
PP	
XX 26-FEB-1999; 99US-0122487.	
PR	
XX (GEST) GENSET.	
PA	
XX Dumas Milne Edwards J, Duclert A, Giordano J;	
PI	
XX WPI; 2000-500381/45.	
XX	
XX	

•

```
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-014732.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156456.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 73.6%; Score 16.2; DB 21; Length 458;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCTCGATCATCT 22
   |||| ||||| |||||
Db 362 TCGACCGGTTTCGGCTCATCT 382

RESULT 14
ABV13083
ID ABV13083 standard; cDNA; 470 BP.
XX
AC ABV13083;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 13074.
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW Pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-0505171.
XX
PR 17-FEB-2000; 2000US-18319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 2160; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV0010-ABV6213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
```

```

CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX SQ Sequence 470 BP; 129 A; 101 C; 118 G; 122 T; 0 other;
Query Match 73.6%; Score 16.2; DB 23; Length 470;
Best Local Similarity 85.7%; Pred. No. 1.le+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 TCGACGGTCTGAGTCATCT 22
Db 444 TGGACGGTCTGAGTCATCT 464
RESULT 15
AAC43639
ID AAC43639 standard; DNA; 911 BP.
XX
AC AAC43639;
XX
DT 17-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 39953.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
XX OS Zea mays subsp. mays.
XX
XX EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140895.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

```

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 73.6%; Score 16.2; DB 21; Length 911;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCTGAGTCATCT 22
Db 370 TCGACCGGTTCCGGGTCATCT 390

Search completed: January 9, 2003, 21:33:51
Job time : 16.4381 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 21:50:31 ; Search time 1.85235 Seconds
(without alignments)
3642.332 Million cell updates/sec

Title: US-09-581-500B-13

Perfect score: 22

Sequence: 1 atcgaaaggtttctgagtcattc 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB ID	Description
	Score	Match	Length		
C 1	17.2	78.2	1589	US-08-335-844A-6	Sequence 6, Appli
C 2	17.2	78.2	3084	US-08-335-844A-19	Sequence 19, Appl
C 3	15.6	70.9	398	US-09-228-986-55	Sequence 55, Appl
C 4	15.6	70.9	461	US-09-228-986-51	Sequence 51, Appl
C 5	15.6	70.9	2447	US-08-387-707-12	Sequence 12, Appl
C 6	15.6	70.9	2447	US-08-405-271A-12	Sequence 12, Appl
C 7	15.2	69.1	1436	US-09-270-542-91	Sequence 91, Appl
C 8	15.2	69.1	1445	US-09-270-542-89	Sequence 89, Appl
C 9	15.2	69.1	1709	US-09-270-542-87	Sequence 87, Appl
C 10	15.2	69.1	2436	US-09-270-542-85	Sequence 85, Appl
C 11	14.8	67.3	802	US-09-724-864-12	Sequence 12, Appl
C 12	14.6	66.4	1248	US-09-134-001C-1980	Sequence 1980, Ap
C 13	14.6	66.4	1378	US-09-149-476-208	Sequence 208, App
C 14	14.6	66.4	1473	US-08-541-033A-25	Sequence 25, Appl
C 15	14.6	66.4	1473	US-08-828-451-25	Sequence 25, Appl
C 16	14.6	66.4	1506	US-08-541-033A-23	Sequence 23, Appl
C 17	14.6	66.4	1506	US-08-828-451-23	Sequence 23, Appl
C 18	14.6	66.4	1740	US-08-991-944-1	Sequence 1, Appli
C 19	14.6	66.4	1969	US-08-541-033A-7	Sequence 7, Appli
C 20	14.6	66.4	1969	US-08-828-451-7	Sequence 7, Appli
C 21	14.6	66.4	2096	US-08-541-033A-19	Sequence 19, Appl
C 22	14.6	66.4	2096	US-08-828-451-19	Sequence 19, Appl
C 23	14.6	66.4	2099	US-08-541-033A-3	Sequence 3, Appli
C 24	14.6	66.4	2099	US-08-828-451-3	Sequence 3, Appli
C 25	14.6	66.4	2137	US-08-541-033A-18	Sequence 18, Appl
C 26	14.6	66.4	2137	US-08-828-451-18	Sequence 18, Appl
C 27	14.6	66.4	2140	US-08-541-033A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-08-335-844A-6/c
; Sequence 6, Application US/08335844A

; Patent No. 6066503

; GENERAL INFORMATION:

; APPLICANT: GRAHAM, MARGARET

; APPLICANT: SMITH, TREVOR STANLEY

; APPLICANT: MUNN, EDWARD ALBERT

; APPLICANT: KNOX, DAVID PATRICK

; APPLICANT: OLIVER, JOANNA JANE

; APPLICANT: NEWTON, SUSAN ELIZABETH

; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING

; TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF

; TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS

; NUMBER OF SEQUENCES: 73

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Rothwell, Figg, Ernst & Kurz

; STREET: Suite 701-E, 555 Thirteenth St., N.W

; CITY: Washington

; STATE: D. C.

; COUNTRY: U.S.A.

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/335,844A

; FILING DATE: 09-JAN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB PCT/GB93/00943

; FILING DATE: 06-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9209936

; FILING DATE: 08-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: WALKER, Barbara W.

; REGISTRATION NUMBER: 35,400

; REFERENCE/DOCKET NUMBER: 1181-223A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)783-6040

; TELEFAX: (202)783-6031

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1689 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

Sequence 1, Appli
Sequence 14, Appli
Sequence 24, Appli
Patent No. 5240838
Sequence 2, Appli
Sequence 204, App
Sequence 5, Appli
Sequence 58, Appli
Sequence 3, Appli
Sequence 7, Appli
Sequence 72, Appli
Sequence 22, Appli
Sequence 22, Appli
Sequence 4, Appli
Sequence 7, Appli
Sequence 2, Appli

MOLECULE TYPE: DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-335-844A-6

Query Match 78.28; Score 17.2; DB 3; Length 1689;
Best Local Similarity 86.48; Pred. No. 5.1; Mismatches 0; Indels 3; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22
||||| ||||| ||||| |||||
DB 505 ATCGAACGGTCTGATACATCT 484

RESULT 2

US-08-335-844A-19/c
Sequence 19, Application US/08335844A
Patent No. 606503

GENERAL INFORMATION:
APPLICANT: GRAHAM, MARGARET
APPLICANT: SMITH, TREVOR STANLEY
APPLICANT: MUNN, EDWARD ALBERT
APPLICANT: KNOX, DAVID PATRICK
APPLICANT: OLIVER, JOANNA JANE
APPLICANT: NEWTON, SUSAN ELIZABETH

TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING
TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF
TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS
NUMBER OF SEQUENCES: 73

CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurz
STREET: Suite 701-E, 555 Thirteenth St., N.W.
CITY: Washington
STATE: D. C.
COUNTRY: U.S.A.
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/335,844A
FILING DATE: 09-JAN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB PCT/GB93/00943
FILING DATE: 06-MAY-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9209936
FILING DATE: 08-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: WALKER, Barbara W.
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 1181-223A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:
LENGTH: 3084 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-335-844A-19

Query Match 78.28; Score 17.2; DB 3; Length 3084;
Best Local Similarity 86.48; Pred. No. 5.6; Mismatches 0; Indels 3; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22
||||| ||||| ||||| |||||
DB 1429 ATCGAACGGTCTGATACATCT 1408

RESULT 3

US-09-228-986-55/c
Sequence 55, Application US/09228986
Patent No. 6359198

GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Niels
TITLE OF INVENTION: Compositions Isolated from Plant Cells
TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
FILE REFERENCE: 11000/1020
CURRENT APPLICATION NUMBER: US/09/228,986
CURRENT FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 130
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 55
LENGTH: 398
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-228-986-55

Query Match 70.9%; Score 15.6; DB 4; Length 398;
Best Local Similarity 81.8%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22
||||| ||||| ||||| |||||
DB 292 ATCGAACGGTCTGATCCACT 271

RESULT 4

US-09-228-986-51/c
Sequence 51, Application US/09228986
Patent No. 6359198

GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Niels
TITLE OF INVENTION: Compositions Isolated from Plant Cells
TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
FILE REFERENCE: 11000/1020
CURRENT APPLICATION NUMBER: US/09/228,986
CURRENT FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 130
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 51
LENGTH: 461
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-228-986-51

Query Match 70.9%; Score 15.6; DB 4; Length 461;
Best Local Similarity 81.8%; Pred. No. 29;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22
||||| ||||| ||||| |||||
DB 237 ATCGAACGGTCTGATCCACT 216

RESULT 5

US-08-387-707-12/c
Sequence 12, Application US/08387707
Patent No. 6265563

GENERAL INFORMATION:
APPLICANT: EVANS, CHRISTOPHER J.
APPLICANT: KEITH, DUANE E.
TITLE OF INVENTION: OPIOID RECEPTOR GENES
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, N.W. Suite 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
- APPLICATION NUMBER: US/08/387,707
FILING DATE: 10-SEP-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22000-20526.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2447 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-387-707-12

Query Match 70.9%; Score 15.6; DB 4; Length 2447;
Best Local Similarity 81.8%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCGACTCATCT 22
||| | ||||| |||||
Db 750 ATCAAGTGGTTCGACTCATCT 729

RESULT 6
US-08-405-271A-12/c
Sequence 12, Application US/08405271A
Patent No. 6432652
GENERAL INFORMATION:
APPLICANT: EVANS, CHRISTOPHER J.
APPLICANT: KEITH, DUANE E.
TITLE OF INVENTION: OPIOID RECEPTOR GENES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 PENNSYLVANIA AVENUE, NW, Suite 5500
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,271A
FILING DATE: 14-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURASHIGE, KATE H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22000-20526.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0763
TELEX: 90-4030 MRSNFOERSWSH
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 2447 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-405-271A-12

Query Match 70.9%; Score 15.6; DB 4; Length 2447;
Best Local Similarity 81.8%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCGACTCATCT 22
||| | ||||| |||||
Db 750 ATCAAGTGGTTCGACTCATCT 729

RESULT 7
US-09-270-542-91/c
Sequence 91, Application US/09270542
Patent No. 6322976
GENERAL INFORMATION:
APPLICANT: Aitman, Timothy
APPLICANT: Scott, James
APPLICANT: Stanton, Lawrence
TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and
FILE REFERENCE: 4198/78179
CURRENT APPLICATION NUMBER: US/09/270,542
CURRENT FILING DATE: 1999-03-17
EARLIER APPLICATION NUMBER: 09/221,222
EARLIER FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 207
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 91
LENGTH: 1436
TYPE: DNA
ORGANISM: Rattus norvegicus
US-09-270-542-91

Query Match 69.1%; Score 15.2; DB 4; Length 1436;
Best Local Similarity 85.0%; Pred. No. 56;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCGAGTCATC 21
||| | ||||| |||||
Db 1104 TCGATAGGTTCTGAGATC 1085

RESULT 8
US-09-270-542-89/c
Sequence 89, Application US/09270542
Patent No. 6322976
GENERAL INFORMATION:
APPLICANT: Aitman, Timothy
APPLICANT: Scott, James
APPLICANT: Stanton, Lawrence
TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and
FILE REFERENCE: 4198/78179
CURRENT APPLICATION NUMBER: US/09/270,542
CURRENT FILING DATE: 1999-03-17
EARLIER APPLICATION NUMBER: 09/221,222
EARLIER FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 207
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 89
LENGTH: 1445
TYPE: DNA
ORGANISM: Rattus norvegicus
US-09-270-542-89

Query Match 69.1%; Score 15.2; DB 4; Length 1445;
Best Local Similarity 85.0%; Pred. No. 56;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTCTCGAGTCATC 21
|||||
Db 1113 TCGATAGGTTCTGAGACATC 1094

RESULT 9
US-09-270-542-87/c
; Sequence 87, Application US/09270542
; Patent No. 6322976
; GENERAL INFORMATION:
; APPLICANT: Altman, Timothy
; APPLICANT: Scott, James
; APPLICANT: Stanton, Lawrence
; TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and
; FILE REFERENCE: 4198/78179
; CURRENT APPLICATION NUMBER: US/09/270,542
; CURRENT FILING DATE: 1999-03-17
; EARLIER APPLICATION NUMBER: 09/221,222
; EARLIER FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 87
; LENGTH: 1709
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (540)..(550)
; OTHER INFORMATION: The N at positions 540, 546, and 550 can be any
; OTHER INFORMATION: nucleotide because the author is unsure of the
; OTHER INFORMATION: exact sequence at these positions.
US-09-270-542-87

Query Match 69.1%; Score 15.2; DB 4; Length 1709;
Best Local Similarity 85.0%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTCTCGAGTCATC 21
|||||
Db 1113 TCGATAGGTTCTGAGACATC 1094

RESULT 10
US-09-270-542-85/c
; Sequence 85, Application US/09270542
; Patent No. 6322976
; GENERAL INFORMATION:
; APPLICANT: Altman, Timothy
; APPLICANT: Scott, James
; APPLICANT: Stanton, Lawrence
; TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and
; FILE REFERENCE: 4198/78179
; CURRENT APPLICATION NUMBER: US/09/270,542
; CURRENT FILING DATE: 1999-03-17
; EARLIER APPLICATION NUMBER: 09/221,222
; EARLIER FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-270-542-85

Query Match 69.1%; Score 15.2; DB 4; Length 2436;
Best Local Similarity 85.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTCTCGAGTCATC 21

Db 1146 TCGATAGGTTCTGAGACATC 1127
|||||
|||||

RESULT 11
US-09-724-864-12/c
; Sequence 12, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
; APPLICANT: Watson, James D
; APPLICANT: Murison, James G.
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed
; TITLE OF INVENTION: by the polynucleotides and methods for their use.
; FILE REFERENCE: 11000.105001
; CURRENT APPLICATION NUMBER: US/09/724,864
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 802
; TYPE: DNA
; ORGANISM: Mouse
US-09-724-864-12

Query Match 67.3%; Score 14.8; DB 4; Length 802;
Best Local Similarity 88.9%; Pred. No. 82;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACGGTCTCGAGTCATCT 22
|||||
Db 399 AACAGGTCCTGAGTCATCT 382

RESULT 12
US-09-134-001C-1980/c
; Sequence 1980, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1980
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1980

Query Match 66.4%; Score 14.6; DB 4; Length 1248;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTCGAGTCATC 21
|||||
Db 1140 ATTAACGGTCTGAGTCATC 1120

RESULT 13
US-09-149-476-208
; Sequence 208, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins

```

/ FILE REFERENCE: pz002p1
/
/ CURRENT APPLICATION NUMBER: US/09/149,476
/ CURRENT FILING DATE: 1998-09-08
/ EARLIER APPLICATION NUMBER: PCT/US98/04493
/ EARLIER FILING DATE: 1998-03-06
/ EARLIER APPLICATION NUMBER: 60/040,162
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,333
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/038,621
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,626
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,334
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,336
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,163
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/047,600
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,615
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,597
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,502
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,633
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,583
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,617
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,618
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,503
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,592
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,581
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,584
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,500
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,587
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,492
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,598
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,613
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,582
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,596
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,612
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,632
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,601
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/043,580
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,568
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,314
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,569
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,311
/ EARLIER FILING DATE: 1997-04-11

```

ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2427 N W 41st Street Suite 2-1

NAME: Whitlock, Ted W.

Search completed: January 10, 2003, 06:37:34
Job time : 3.85235 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 01:01:36 ; Search time 2.87949 Seconds
(without alignments)
3362.497 Million cell updates/sec

Title: US-09-581-500B-13

Perfect score: 22
Sequence: 1 atcgaacggtcttgatcatct 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	17.2	78.2	846	9	US-09-938-842A-421
2	17.2	78.2	1889	10	US-09-775-879-20
3	17.2	78.2	2925	10	US-09-775-879-22
4	15.8	71.8	1293	10	US-09-974-300-30
5	15.6	70.9	292	10	US-09-294-093B-461
6	15.6	70.9	2313	9	US-09-938-842A-2640
7	15.6	70.9	2447	10	US-09-823-114-12
8	15.6	70.9	3109	10	US-09-925-301-544
9	15.6	70.9	3954	9	US-10-071-766-44
10	15.4	70.0	4006	10	US-09-925-300-580
11	15.2	69.1	1469	10	US-09-917-800A-499
12	15.2	69.1	2484	10	US-09-747-835A-5
13	15.2	69.1	3188	10	US-09-747-835A-3
14	15.2	69.1	21252	10	US-09-070-927A-94
15	15.2	69.1	23934	10	US-09-764-860-777
16	15.2	69.1	23934	10	US-09-764-877-2336
17	15.2	69.1	23934	10	US-09-764-877-2544
18	15.2	69.1	70768	9	US-10-135-322-13
19	15	68.2	510	10	US-09-764-877-146

20	15	68.2	991	10	US-09-764-877-2302
21	15	68.2	1868	10	US-09-764-877-2301
22	14.8	67.3	990	9	US-09-938-842A-155
23	14.6	66.4	258	10	US-09-923-876-2317
24	14.6	66.4	422	10	US-09-924-035A-476
25	14.6	66.4	610	9	US-09-736-457-1094
26	14.6	66.4	610	9	US-09-902-941-1094
27	14.6	66.4	610	9	US-09-849-626-1094
28	14.6	66.4	655	10	US-09-770-149-502
29	14.6	66.4	687	10	US-09-917-800A-1219
30	14.6	66.4	693	9	US-09-938-842A-1871
31	14.6	66.4	816	10	US-09-815-242-9194
32	14.6	66.4	1077	9	US-09-938-842A-1129
33	14.6	66.4	1189	10	US-09-900-237-5
34	14.6	66.4	1473	10	US-09-070-844-25
35	14.6	66.4	1506	10	US-09-070-844-23
36	14.6	66.4	1583	12	US-10-062-254-273
37	14.6	66.4	1686	9	US-09-938-842A-564
38	14.6	66.4	1969	10	US-09-070-844-7
39	14.6	66.4	2000	9	US-09-938-842A-4819
40	14.6	66.4	2096	10	US-09-070-844-19
41	14.6	66.4	2099	10	US-09-070-844-3
42	14.6	66.4	2137	10	US-09-070-844-18
43	14.6	66.4	2140	10	US-09-070-844-1
44	14.6	66.4	2319	9	US-10-121-032-14
45	14.6	66.4	2733	9	US-09-935-868-27

ALIGNMENTS

RESULT 1
US-09-938-842A-421
; Sequence 421, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepes, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938, 842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227, 866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264, 647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300, 111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 421
; LENGTH: 846
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-421

Query Match 78.2%; Score 17.2; DB 9; Length 846;
Best Local Similarity 86.4%; Pred. No. 8.1;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTCGATCATCT 22
Db 56 ACCGAATCGTCTCGATCATCT 77

RESULT 2
US-09-775-879-20
; Sequence 20, Application US/09775879
; Patent No. US2002006822A1
; GENERAL INFORMATION:
; APPLICANT: Choe, Sunghwa

```
; APPLICANT: Feldmann, Kenneth A
; TITLE OF INVENTION: Dwf7 MUTANTS
; FILE REFERENCE: 2225-0003
; CURRENT APPLICATION NUMBER: US/09/775,879
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/179,901
; PRIOR FILING DATE: 2000-02-02
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 1889
; TYPE: DNA
; ORGANISM: Genomic dwf7 (Arabidopsis)
US-09-775-879-20

Query Match      78.2%; Score 17.2; DB 10; Length 1889;
Best Local Similarity 86.4%; Pred. No. 9.1;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCTGAGTCATCT 22
Db 198 ACCGAATCGTTCTGAGTCATCT 219

RESULT 3
US-09-775-879-22
; Sequence 22, Application US/09775879
; Patent No. US2002006822A1
; GENERAL INFORMATION:
; APPLICANT: Choe, Sunghwa
; APPLICANT: Feldmann, Kenneth A
; TITLE OF INVENTION: Dwf7 MUTANTS
; FILE REFERENCE: 2225-0003
; CURRENT APPLICATION NUMBER: US/09/775,879
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/179,901
; PRIOR FILING DATE: 2000-02-02
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 2925
; TYPE: DNA
; ORGANISM: Genomic HDF7
US-09-775-879-22

Query Match      78.2%; Score 17.2; DB 10; Length 2925;
Best Local Similarity 86.4%; Pred. No. 9.8;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCTGAGTCATCT 22
Db 1564 ACCGAATGGTTCTGAGTCACCT 1585

RESULT 4
US-09-974-300-30
; Sequence 30, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30

; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 461
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700342780H1
; NAME/KEY: unsure
; LOCATION: 266-267
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-461

Query Match      70.9%; Score 15.6; DB 10; Length 292;
Best Local Similarity 81.8%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCTGAGTCATCT 22
Db 254 ATCAACGGGTTTGAGTAATCT 233

RESULT 6
US-09-938-842A-2640
; Sequence 2640, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2640
```

```
; LENGTH: 2313
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2640

Query Match      70.9%; Score 15.6; DB 9; Length 2313;
Best Local Similarity 81.8%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAAGGTTCTGAGTCATCT 22
    ||| || ||||| ||||| |||
Db 411 ATCGAAGGTTCTGAGTCAGCT 432

RESULT 7
US-09-823-114-12/C
; Sequence 12, Application US/09823114
; Patent No. US20020061554A1
; GENERAL INFORMATION:
; APPLICANT: EVANS, CHRISTOPHER J.
; KEITH, DUANE E.
; TITLE OF INVENTION: OPIOID RECEPTOR GENES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 PENNSYLVANIA AVENUE, NW, Suite 5500
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/823,114
; FILING DATE: 29-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/148,351
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 22000-20526.22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-823-114-12

Query Match      70.9%; Score 15.6; DB 10; Length 2447;
Best Local Similarity 81.8%; Pred. No. 64;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAAGGTTCTGAGTCATCT 22
    ||| | ||||| ||||| |||
Db 750 ATCAAGTGGTTCTGACTCATCT 729

RESULT 8
US-09-925-301-544
; Sequence 544, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCR/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 544
; LENGTH: 3109
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1011)
; OTHER INFORMATION: n equals a,t,g; or c
US-09-925-301-544

Query Match      70.9%; Score 15.6; DB 10; Length 3109;
Best Local Similarity 81.8%; Pred. No. 66;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAAGGTTCTGAGTCATCT 22
    ||||| ||||| ||||| |||
Db 386 ATCGAAGCTTCTGAATCATTT 407

RESULT 9
US-10-071-766-44
; Sequence 44, Application US/10071766
; Publication No. US20020192678A1
; GENERAL INFORMATION:
; APPLICANT: Hwei-Mei Chen
; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE
; FILE REFERENCE: PA-0043 US
; CURRENT APPLICATION NUMBER: US/10/071,766
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: PERL Program
; SEQ ID NO 44
; LENGTH: 3954
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020192678A1 251859.2
; LOCATION: 3786, 3788, 3791
; OTHER INFORMATION: a, t, c, g, or other
US-10-071-766-44

Query Match      70.9%; Score 15.6; DB 9; Length 3954;
Best Local Similarity 81.8%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAAGGTTCTGAGTCATCT 22
    ||||| ||||| ||||| |||
Db 1157 ATCGAAGCTTCTGAATCATTT 1178

RESULT 10
US-09-925-300-580
; Sequence 580, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
```



```

; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 580
; LENGTH: 4006
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-580

Query Match      70.0%; Score 15.4; DB 10; Length 4006;
Best Local Similarity 94.1%; Pred. No. 87;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 ACGGTTCTGAGTCATCT 22
    |||||
Db 237 ACGGTTCTGAGTCACCT 253

RESULT 11
US-09-917-800A-499/c
; Sequence 499, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 499
; LENGTH: 1469
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AF072411
US-09-917-800A-499

Query Match      69.1%; Score 15.2; DB 10; Length 1469;
Best Local Similarity 85.0%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCTGAGTCATC 21
    |||||
Db 1097 TCGATAGGTTCTGAGATC 1078

; Sequence 5, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Asundi, Vinod
; APPLICANT: Dmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-
; TITLE OF INVENTION: LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: HYS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747,835A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 2484
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-747-835A-5

Query Match      69.1%; Score 15.2; DB 10; Length 2484;
Best Local Similarity 85.0%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CGAACGGTTCTGAGTCATCT 22
    |||
Db 1875 CGATGGTTCTGGGTCACT 1894

RESULT 13
US-09-747-835A-3
; Sequence 3, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Asundi, Vinod
; APPLICANT: Dmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-
; TITLE OF INVENTION: LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: HYS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747,835A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312

```

```
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 3188
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2484)
US-09-747-835A-3

Query Match          59.1%; Score 15.2; DB 10; Length 3188;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 CGAACGGTTCGAGTCATCT 22
      ||| ||||| |||||
Db      1875 CGATGGTCTGGGTCATCT 1894

RESULT 14
US-09-070-927A-94
; Sequence 94, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
;          Patrick J. Dillon
;          Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21252 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

```
; SEQUENCE DESCRIPTION: SEQ ID NO: 94:
US-09-070-927A-94

Query Match          69.1%; Score 15.2; DB 10; Length 21252;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 TCGAACGGTTCGAGTCATC 21
      ||| ||||| |||||
Db      10687 TCAAACGTCTCTGATTCATC 10706

RESULT 15
US-09-764-860-777/c
; Sequence 777, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 777
; LENGTH: 23934
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-860-777

Query Match          69.1%; Score 15.2; DB 10; Length 23934;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ATCGAACGGTTCGAGTCAT 20
      ||| ||||| |||||
Db      13230 ATAAAGGGTTCGAGTCAT 13211

Search completed: January 10, 2003, 06:43:17
Job time : 8.87949 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 21:14:32 ; Search time 74.5156 Seconds
(without alignments)
4781.561 Million cell updates/sec

Title: US-09-581-500B-13

Perfect score: 22

Sequence: 1 atcgaaaggttgtgatcatct 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estinv:*
4: em_estpln:*
5: em_estvrt:*
6: em_estfun:*
7: em_estmam:*
8: em_estmus:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_estfun:*
15: em_estfun:*
16: em_estmam:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	244	9	AA889105 am38h04.s
2	22	100.0	270	9	AA904435 ok07g11.s
3	22	100.0	636	14	BQ807824 NISC.kk10
4	22	100.0	930	14	BQ435307 AGENCOURT
5	19.4	88.2	408	12	BG452389 NF097F02L
6	19.4	88.2	529	12	BE999204 EST430927

c 7	19.4	88.2	632	12	BF639518
c 8	19.4	88.2	653	10	BE318984
c 9	19.4	88.2	668	12	BG449328
c 10	17.8	80.9	715	12	BG467650
c 11	17.8	80.9	773	17	BH590723
c 12	17.2	78.2	143	17	AQ957346
c 13	17.2	78.2	298	12	BF899987
c 14	17.2	78.2	349	12	BF899989
c 15	17.2	78.2	396	12	BF738194
c 16	17.2	78.2	424	10	AW374777
c 17	17.2	78.2	425	12	BE932180
c 18	17.2	78.2	433	12	BF738261
c 19	17.2	78.2	447	12	BG008704
c 20	17.2	78.2	450	12	BG008684
c 21	17.2	78.2	456	17	AQ914000
c 22	17.2	78.2	458	10	AW062933
c 23	17.2	78.2	459	10	AW003072
c 24	17.2	78.2	459	10	AW962224
c 25	17.2	78.2	460	10	AW266225
c 26	17.2	78.2	460	10	AW374720
c 27	17.2	78.2	475	10	AW062934
c 28	17.2	78.2	489	10	AW374725
c 29	17.2	78.2	499	12	BF883208
c 30	17.2	78.2	509	17	AQ739516
c 31	17.2	78.2	544	9	AI670991
c 32	17.2	78.2	550	9	AJ285111
c 33	17.2	78.2	558	17	BH387822
c 34	17.2	78.2	598	17	BH383813
c 35	17.2	78.2	618	10	AW375534
c 36	17.2	78.2	669	14	BF002543
c 37	17.2	78.2	698	10	AW994568
c 38	17.2	78.2	701	12	BF295216
c 39	17.2	78.2	731	17	AQ258221
c 40	17.2	78.2	744	10	AW053439
c 41	17.2	78.2	749	17	CNS01E3P
c 42	17.2	78.2	782	12	BF529413
c 43	17.2	78.2	858	17	CNS01L30
c 44	17.2	78.2	901	12	BF964582
c 45	17.2	78.2	927	17	CNS01NOK

ALIGNMENTS

RESULT 1
AA889105
LOCUS am38h04.sl Soares_NFL_T_GBC_Sl Homo sapiens cDNA clone
DEFINITION IMAGE:1471063 3', mRNA sequence.
ACCESSION AA889105
VERSION AA889105.1 GI:3015984
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 244)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)
JOURNAL Tumor Gene Index
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 482 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 192.
Location/Qualifiers
1..244

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1471063"

BF639518 NF012C12I
BE318984 NF043F02L
BG449328 NF043G05I
BG467650 Na.L3.01D
BH590723 BOHIE73TF
AQ957346 LERAP27TF
BF899987 CM3-MT019
BF899989 CM3-MT019
BF738194 CM3-KT003
AW374777 MR1-CT005
BE932180 IL3-HT061
BF738261 CM3-KT003
BG008704 RC2-GN026
BG008684 RC2-GN026
AQ914000 nbe00046J
AW062933 MR0-ST003
AW003072 wt03b04.x
AW962224 EST374297
AW266225 L30-2934T
AW374720 MR1-CT005
AW062934 MR0-ST003
AW374725 MR1-CT005
BF883208 QV3-ET017
AQ739516 HS_4506_A
AI670991 WD12F09.x
AJ285111 4A3B-AAP-
BH387822 AG-ND-130
BH383813 AG-ND-137
AW375534 QV0-CT018
BF002543 BF002543
AW994568 RC0-BN003
BF295216 014PBA02
AQ258221 nbxb0019N
AW053439 L30-1352T
AL140234 Anopheles
BF529413 602041936
AL149141 Anopheles
BF964582 602267886
AL152485 Anopheles

/clone_lib="Soares_NFL_T_GBC_S1"

/lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site.1: Not I; Site.2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI-CGAP_GCB1) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 74 a 61 c 63 g 46 t

Query Match 100.0%; Score 22; DB 9; Length 244;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTTCGAGTCATCT 22
 |||||

Db 121 ATCGACGGTTCGAGTCATCT 142

RESULT 2
 AA904435

LOCUS ok07q11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:1507172 3', mRNA sequence.

ACCESSION AA904435

VERSION AA904435.1 GI:3039558

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 270)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL Tumor Gene Index

COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 437 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 232.

Location/Qualifiers

1..270

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1507172"

/clone_lib="Soares_NFL_T_GBC_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site.1: Not I; Site.2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBHL19W, testis NHT, and B-cell

NCI-CGAP_GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo."

77 a 66 c 74 g 53 t

BASE COUNT

ORIGIN

Query Match

100.0%; Score 22; DB 9; Length 270;

Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTTCGAGTCATCT 22
 |||||

Db 121 ATCGACGGTTCGAGTCATCT 142

RESULT 3

BO807824/c

LOCUS

DEFINITION NISC.k10f04.y1 NCI-CGAP_Brn72 Macaca mulatta cDNA clone
 IMAGE:5331199 5', mRNA sequence.

ACCESSION BO807824

VERSION BO807824.1 GI:22032033

KEYWORDS EST.

SOURCE rhesus monkey.

ORGANISM

Macaca mulatta

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
 Cercopitheidae; Macaca.

REFERENCE 1 (bases 1 to 636)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL Tumor Gene Index

COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

cDNA Library Preparation:

DNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC)

Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 info@image.llnl.gov

Plate: LLAM1840 row: K column: 8

Seq primer: M13Rpl reverse primer (ABI).

Location/Qualifiers

1..636

FEATURES

source

/organism="Macaca mulatta"

/db_xref="taxon:9544"

/clone="IMAGE:5331199"

/clone_lib="NCI-CGAP_Brn72"

/tissue_type="hypothalamus"

/note="Organ: brain; Vector: pCMV-Sport6.cdb; Site.1:

NotI; Site.2: EcoRV; Cloned unidirectionally. Primer:

Oligo dT. Average insert size 2.2 kb. Constructed by

Invitrogen. Note: this is a NCI-CGAP Library."

81 a 210 c 190 g 155 t

BASE COUNT

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 636;

Best Local Similarity 100.0%; Pred. No. 5.4;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTTCGAGTCATCT 22
 |||||

Db 431 ATCGACGGTTCGAGTCATCT 410

RESULT 4

BO435307

LOCUS

DEFINITION BO435307 930 bp mRNA linear EST 24-MAY-2002
 AGENCOURT_7926752 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6160773
 5', mRNA sequence.

ACCESSION BO435307

VERSION BO435307.1 GI:21174383

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 930)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13511 row: p column: 22
High quality sequence stop: 607.
FEATURES             source
1..930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6160773"
/clone_lib="NIH.MGC.72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: skin; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 268 a 210 c 263 g 186 t 3 others
ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 930;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGAAGCGTTCGAGTCATCT 22
|||||
DB 36 ATCGAAGCGTTCGAGTCATCT 57
|||||

RESULT 5
BG452389/c
LOCUS BG452389 408 bp mRNA linear EST 16-MAR-2001
DEFINITION NF097F02LIF1016 Developing leaf Medicago truncatula cDNA clone
ACCESSION BG452389
VERSION BG452389.1 GI:13371183
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 408)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Iman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula leaf library
Unpublished (2000)
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert Length: 408 Std Error: 0.00
Plate: 097 row: F column: 02
Seq primer: TCACACGAGGACACTATGAC.
Location/Qualifiers
1..408
/organism="Medicago truncatula"
/db_xref="taxon:3880"

```

```

/clone="NF097F02LIF"
/clone_lib="Developing leaf"
/tissue_type="leaf"
/dev_stage="Pooled developmental"
/Note="Vector: Lambda Zap; Contains a mixture of very
young, developing, mature and senescing leaves."
BASE COUNT 98 a 114 c 94 g 99 t 3 others
ORIGIN

Query Match 88.2%; Score 19.4; DB 12; Length 408;
Best Local Similarity 95.2%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCGAGTCATCT 22
|||||
DB 242 TCGAACGGTTCGAGTCATCT 222
|||||

RESULT 6
BE999204/c
LOCUS BE999204 529 bp mRNA linear EST 06-OCT-2000
DEFINITION EST430927 GVSN Medicago truncatula cDNA clone pgVSN-15B23, mRNA
SEQUENCE.
ACCESSION BE999204
VERSION BE999204.1 GI:10699480
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 529)
Fedorova, M., Pierson, B.L., Samac, D.A., Gantt, J.S., Vance, C.P., Town
C.D., Bowman, C.L., Craven, M.B., Cho, J. and Fraser, C.M.
ESTs from senescent nodules of Medicago truncatula
Unpublished (2000)
Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
University of Minnesota name: M273182e TIGR sequence name:
MTKBA127K More information is available at:
http://chrysie.tamu.edu/medicago
Seq primer: SKnod (CTA gAA CTA gTg gAT CC).
Location/Qualifiers
1..529
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pgVSN-15B23"
/clone_lib="GVSN"
/tissue_type="senescent root nodules"
/dev_stage="mixture of effective nodules from 40 day old
plants harvested 36 hours post shoot removal and nodules
collected from 2 month old plants at mid-pod stage"
/lab_host="E. coli strain SOLR"
/Note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from the
mixture of effective nodules of 40 day old plants
harvested 36 hours post shoot removal and nodules
collected from 2 month old plants at mid-pod stage. The
cDNA was directionally ligated into the Uni-ZAP XR vector
from stragene and packaged using Gigapack III Gold
packaging extracts. Plasmids containing cDNA inserts were
excised from the recombinant lambda-ZAP phage using
Ex-Assist helper phage and propagated in SOLR cells."
BASE COUNT 122 a 119 c 145 g 143 t
ORIGIN

```

Query Match 88.2%; Score 19.4; DB 12; Length 529;
 Best Local Similarity 95.2%; Pred. No. 85;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGAACGGTCTGAGTCATCT 22
 DB 69 TCGAACGGTCTGAGTCATCT 49

RESULT 7
 BF639518/c 632 bp mRNA linear EST 19-DEC-2000
 LOCUS NF012C12IN1087 Insect herbivory Medicago truncatula cDNA clone
 DEFINITION NF012C12IN 5', mRNA sequence.
 ACCESSION BF639518
 VERSION BF639518.1 GI:11903676
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.

REFERENCE 1 (bases 1 to 632)
 Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
 ,H.R., Inman,J.F., Weller,J.W. and May,G.D.
 Expressed Sequence Tags from the Samuel Roberts Noble Foundation
 Medicago truncatula insect herbivory library
 UNPUBLISHED (2000)
 JOURNAL Contact: Korth K
 COMMENT Dept. of Plant Pathology
 University of Arkansas
 217 Plant Science Building, Fayetteville, AR 72701, USA
 Tel: 501 575 5191
 Fax: 501 575 7601
 Email: kkorth@comp.uark.edu
 Insert Length: 632 Std Error: 0.00
 Plate: 012 row: C column: 12
 Seq primer: TCACACGAGGAACAGCTATGAC.
 FEATURES Location/Qualifiers
 source 1..632
 /organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone_lib="NF012C12IN"
 /clone_lib="Insect herbivory"
 /tissue_type="local and systemic leaves"
 /dev_stage="mature"
 /note="Vector: Lambda Zap; Library was produced from fully
 expanded M. truncatula leaves of plants fed upon by
 Spodoptera exigua (beet armyworm) for 24 hours. Systemic
 (undamaged leaves from injured plants) and wounded leaves
 were harvested and pooled."

BASE COUNT 152 a 163 c 158 g 159 t
 ORIGIN

Query Match 88.2%; Score 19.4; DB 12; Length 532;
 Best Local Similarity 95.2%; Pred. No. 91;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGAACGGTCTGAGTCATCT 22
 DB 223 TCGAACGGTCTGAGTCATCT 203

RESULT 8
 BE318984/c 653 bp mRNA linear EST 21-DEC-2000
 LOCUS NF043F02LIF1016 Developing leaf Medicago truncatula cDNA clone
 DEFINITION NF043F02LIF 5', mRNA sequence.
 ACCESSION BE318984
 VERSION BE318984.2 GI:11959620
 KEYWORDS EST.
 SOURCE barrel medic.

ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.

REFERENCE 1 (bases 1 to 653)
 Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
 Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
 Expressed Sequence Tags from the Samuel Roberts Noble Foundation
 Medicago truncatula leaf library
 UNPUBLISHED (2000)
 JOURNAL Contact: May GD
 COMMENT On Jul 14, 2000 this sequence version replaced gi:9192761.
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA
 Tel: 580 221 7391
 Fax: 580 221 7380
 Email: gmay@noble.org
 Medicago Genome Initiative accession: MGI:S:27143
 Insert Length: 677 Std Error: 0.00
 Plate: 043 row: F column: 02
 Seq primer: TCACACGAGGAACAGCTATGAC.
 FEATURES Location/Qualifiers
 source 1..653
 /organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone_lib="NF043F02LF"
 /clone_lib="Developing leaf"
 /tissue_type="leaf"
 /dev_stage="Pooled developmental"
 /note="Vector: Lambda Zap; Contains a mixture of very
 young, developing, mature and senescing leaves."

BASE COUNT 158 a 171 c 155 g 169 t
 ORIGIN

Query Match 88.2%; Score 19.4; DB 10; Length 653;
 Best Local Similarity 95.2%; Pred. No. 92;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGAACGGTCTGAGTCATCT 22
 DB 259 TCGAACGGTCTGAGTCATCT 239

RESULT 9
 BG449328/c 668 bp mRNA linear EST 16-MAR-2001
 LOCUS NF043G05IN1F1038 Insect herbivory Medicago truncatula cDNA clone
 DEFINITION NF043G05IN 5', mRNA sequence.
 ACCESSION BG449328
 VERSION BG449328.1 GI:13368109
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.

REFERENCE 1 (bases 1 to 668)
 Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
 ,H.R., Inman,J.T., Weller,J.W. and May,G.D.
 Expressed Sequence Tags from the Samuel Roberts Noble Foundation
 Medicago truncatula insect herbivory library
 UNPUBLISHED (2000)
 JOURNAL Contact: Korth K
 COMMENT Dept. of Plant Pathology
 University of Arkansas
 217 Plant Science Building, Fayetteville, AR 72701, USA
 Tel: 501 575 5191
 Fax: 501 575 7601
 Email: kkorth@comp.uark.edu
 Insert Length: 668 Std Error: 0.00

Plate: 043 row: G column: 05
Seq primer: TCACACAGGAAACAGCTATGAC.

FEATURES

source
1. .668
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF043G05IN"
/clone_lib="insect herbivory"
/tissue_type="local and systemic leaves"
/dev_stage="mature"
/note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

BASE COUNT 161 a 168 c 160 g 172 t 7 others

ORIGIN

Query Match 88.2%; Score 19.4; DB 12; Length 668;
Best Local Similarity 95.2%; Pred. No. 93;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGAACGGTCTCGATCATCT 22

||||| ||||| ||||| |||||

Db 223 TCGAACGGTCTCGATCATCT 203

RESULT 10

BG467650/c

LOCUS

DEFINITION Na_L3_01D05_SAC Necator americanus (parasitic nematode) L3 Necator americanus CDNA clone Na_L3_01D05 5', mRNA sequence.

ACCESSION

BG467650

KEYWORDS

EST.

SOURCE

ORGANISM Necator americanus.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida; Ancylostomatidae; Ancylostomatidae; Bunostominae; Necator.

REFERENCE

1 (bases 1 to 715)

AUTHORS

Blaxter, M.L., Parkinson, J., Whitton, C., Daub, J., Guiliano, D., Hall

TITLE

Edinburgh University/Sanger Centre Nematode EST Project

JOURNAL

Unpublished (2000)

COMMENT

Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JF, UK.

Tel: +44 131 650 6760

Fax: +44 131 670 5450

Email: mark.blaxter@ed.ac.uk

The library was prepared for Dr. David Pritchard University of

Edinburgh

PCR Primers

FORWARD: SAC

BACKWARD: T7PL

Plate: 01 row: D column: 05

Seq primer: SAC

High quality sequence stop: 463.

FEATURES

source

1. .715

/organism="Necator americanus"

/db_xref="taxon:51031"

/clone="Na_L3_01D05"

/clone_lib="Necator americanus (parasitic nematode) L3"

/sex="Mixed"

/dev_stage="L3"

/note="Vector: pCMV-PCR vector; Site_1: EcoRI (5'end);

Site_2: XhoI (3'end); Necator americanus is a human

hookworm, responsible for debilitating anaemia. The

library was constructed by David Pritchard (University of

Nottingham, UK) from mRNA from N.americanus L3 stage (Nottingham strain) maintained in hamsters."

BASE COUNT 213 a 150 c 168 g 184 t

ORIGIN

Query Match 80.9%; Score 17.8; DB 12; Length 715;

Best Local Similarity 90.5%; Pred. No. 5.5e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTCGATCATCT 21

||||| ||||| ||||| |||||

Db 483 ATCGAACGGTCTCGATCATCT 463

RESULT 11

BH590723

LOCUS

DEFINITION BOHIE73TF BOHI Brassica oleracea genomic clone BOHIE73, DNA

ACCESSION

BH590723

VERSION

BH590723.1

KEYWORDS

GSS.

SOURCE

ORGANISM Brassica oleracea.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE

1 (bases 1 to 773)

AUTHORS

Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

TITLE

Whole genome shotgun sequencing of Brassica oleracea

JOURNAL

Unpublished (2001)

COMMENT

Other_GSSs: BOHIE73TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: ctown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1. .773

/organism="Brassica oleracea"

/strain="T01000DH3"

/db_xref="taxon:3712"

/clone="BOHIE73"

/clone_lib="BOHI"

/note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared

genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 198 a 209 c 165 g 201 t

ORIGIN

Query Match 80.9%; Score 17.8; DB 17; Length 773;

Best Local Similarity 90.5%; Pred. No. 5.7e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCGAACGGTCTCGATCATCT 22

||||| ||||| ||||| |||||

Db 719 TCGAACCATCTCGATCATCT 739

RESULT 12

AQ957346/c

LOCUS

DEFINITION LERAP27TF LERA Arabidopsis thaliana genomic clone LERAP27, DNA

sequence.

ACCESSION AQ957346

VERSION AQ957346.1

KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

1 (bases 1 to 143)

REFERENCE
AUTHORS Buell, C.R., Lin, X., Pei, G., Barnstead, M., Bowman, C., Utterbach, T., Feldblyum, T., Liang, F., Creasy, T., and Fraser, C.M.
TITLE Genomic survey sequencing of *landsberg erecta* ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms
JOURNAL Unpublished (2000)
COMMENT Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: at@tigr.org
 For additional information, see <http://www.tigr.org/tdb/at/at.html>
 Seq primer: TF
 Class: Shotgun.

FEATURES
 source
 1..143
 /organism="Arabidopsis thaliana"
 /strain="Landsberg erecta"
 /db_xref="taxon:3702"
 /clones="LERAP27"
 /clone_lib="LERA"
 /note="Organ: leaf; Vector: PHOS1; Total genomic DNA was sheared to 0.9-1 Kbp before ligation."

BASE COUNT
 47 a 24 c 15 g 57 t

ORIGIN

Query Match 78.2%; Score 17.2; DB 17; Length 143;
 Best Local Similarity 86.4%; Pred. No. 5.4e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22
 ||||| ||| ||||| |||

DB 64 ATCGACAGTTATGAGTCACT 43

RESULT 13
 BF899987
 LOCUS BF899987 298 bp mRNA linear EST 18-JAN-2001
 DEFINITION CM3-MT0195-051200-525-f04 MT0195 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF899987
 VERSION BF899987.1 GI:12291446
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 298)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-MT0195-051200-525-f04&t3=2000-12-05&t4=1>)

Seq primer: puc 18 forward
 High quality sequence start: 2
 High quality sequence stop: 129.

FEATURES
 source
 1..298
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MT0195"
 /dev_stage="Adult"
 /note="Organ: marrow; Vector: puc18; Site:1; SmaI; Site:2; SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT
 77 a 77 c 66 g 78 t

ORIGIN

Query Match 78.2%; Score 17.2; DB 12; Length 298;
 Best Local Similarity 86.4%; Pred. No. 7.4e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22
 ||||| ||| ||||| |||

DB 156 ACCGACAGTTCTGAGTCATTT 177

RESULT 14
 BF899989/c
 LOCUS BF899989 349 bp mRNA linear EST 18-JAN-2001
 DEFINITION CM3-MT0195-051200-525-h10 MT0195 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF899989
 VERSION BF899989.1 GI:12291448
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 349)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-MT0195-051200-525-h10&t3=2000-12-05&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 19
 High quality sequence stop: 347.

FEATURES
 source
 1..349
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MT0195"
 /dev_stage="Adult"
 /note="Organ: marrow; Vector: puc18; Site:1; SmaI; Site:2; SmaI; A mini-library was made by cloning products derived

